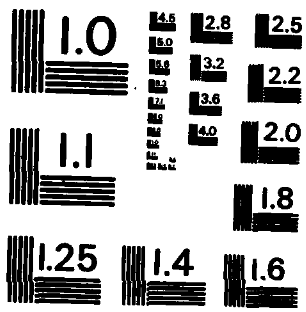
AD-A120 739 ROBUST PROJECTION PURSUIT ESTIMATOR FOR DISPERSION MATRICES AND PRINCIPAL COMPONENTS(U) HARVARD UNIV CAMBRIDGE MASS DEPT OF STATISTICS G LI ET AL. NOV 81 1/1

UNCLASSIFIED PJH-11 N00014-79-C-0512

END



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

RESEARCH REPORT

PJH - 11

ROBUST PROJECTION PURSUIT ESTIMATOR FOR DISPERSION MATRICES AND PRINCIPAL COMPONENTS

Guoying Li and Zhongliang Chen

November 1981

This work was facilitated in part by National Science Foundation Grant MCS-79-08685 and Office of Naval Research Contract N00014-79-C-0512.

Department of Statistics
Harvard University
Cambridge, Massachusetts

ROBUST PROJECTION PURSUIT ESTIMATOR
FOR DISPERSION MATRICES AND PRINCIPAL COMPONENTS

G. Li and Z. Chen*
Department of Statistics, Harvard University

ABSTRACT

This paper proposes and discusses the ROBUST PROJECTION PURSUIT ESTIMATOR for dispersion matrices and their principal components. This estimator finds robust principal components by searching, successively, for directions which maximize (minimize) a robust estimate of scale; the estimate of the dispersion matrix is constructed from the estimated principal components.

These estimators are shown below (under mild conditions) to have a number of desirable properties. They are orthogonally equivariant and, within any elliptic underlying density family, asymptotically affinely equivariant. Furthermore, at elliptic densities, they are consistent and weakly continuous (i.e., qualitatively robust). Finally, they have good quantitative robustness — their breakdown point can be as high as 1/2.

The robust projection pursuit approach is a promising alternative to other estimators of dispersion matrices.

---

*The authors are on leave from the Institute of Systems Science, Academia Sinica, Beijing, People's Republic of China.

ROBUST PROJECTION PURSUIT ESTIMATOR FOR DISPERSION MATRICES AND PRINCIPAL COMPONENTS

G. Li and Z. Chen

1. Introduction

Covariance and correlation matrices and their principal components are basic elements in multivariate data analysis. Yet, the classical estimates of these quantities are very sensitive to outliers. In recent years, several robust alternatives have been proposed and studied (see [1], [3], [4], [8], [9], [10], [14], [15], and [16]). The problem seems to be that no one is really satisfactory. The matrix element methods are not orthogonally equivariant, do not even guarantee to produce a positive definite matrix, and are very difficult to study theoretically; while affinely equivariant M-estimates of covariance matrix have quite poor breakdown properties in high dimension.

This paper, together with [2], proposes and discusses a new type of estimator for covariance/correlation matrices and their principal components. It uses a Projection Pursuit (PP) procedure (see [2], [5], [6], [7], [11], [12]) with a robust estimate of scale as the projection index. So we call it the ROBUST PP-ESTIMATOR. Its idea was originally raised by Huber (see [10], p. 200 and pp. 203-204).

The PP procedure deals with high dimensional data: it searches low-dimensional projections which maximize (minimize) an objective function called PROJECTION INDEX.

Suppose that $X \sim F(\underset{\sim}{x})$ is a p-dimensional random vector and that its location is known in advance and fixed at $\underset{\sim}{0}$. Our robust PP-estimator works as follows: Let $S(\cdot)$ be a robust estimator for scale, which is usually weakly continuous, and $\underset{\sim}{a} \in R^p$ be a vector. Denote the distribution function of $\underset{\sim}{a}^T\underset{\sim}{X}$ by $F^{\underset{\sim}{a}}$, or $\mathcal{F}(\underset{\sim}{a}^T\underset{\sim}{X})$ when needed. The estimators for principal components, in terms of functionals and denoted by $S_i(F)$ and $A_i(F)$ $(1 \le i \le p)$, are defined, either using the maximizing procedure:

$$
\begin{aligned}
&S_1(F) = \max_{|\underset{\sim}{a}|=1} S(F^{\underset{\sim}{a}}) \\
&A_1(F) = \underset{\sim}{a}_1 \quad \text{where} \quad |\underset{\sim}{a}_1| = 1,\ S(F^{\underset{\sim}{a}_1}) = S_1(F)\ , \\
&S_2(F) = \max_{|\underset{\sim}{a}|=1,\ \underset{\sim}{a}\perp\underset{\sim}{a}_1} S(F^{\underset{\sim}{a}})\ , \\
&A_2(F) = \underset{\sim}{a}_2 \quad \text{where} \quad |\underset{\sim}{a}_2| = 1,\ \underset{\sim}{a}_2 \perp \underset{\sim}{a}_1,\ S(F^{\underset{\sim}{a}_2}) = S_2(F)\ , \\
&\cdots\cdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (1.1) \\
&S_{p-1}(F) = \max_{\substack{|\underset{\sim}{a}|=1 \\ \underset{\sim}{a}\perp\underset{\sim}{a}_1,\ldots,\underset{\sim}{a}_{p-2}}} S(F^{\underset{\sim}{a}})\ , \\
&A_{p-1}(F) = \underset{\sim}{a}_{p-1} \quad \text{where} \quad |\underset{\sim}{a}_{p-1}| = 1,\ \underset{\sim}{a}_{p-1} \perp \underset{\sim}{a}_1,\ldots,\underset{\sim}{a}_{p-2},\ S(F^{\underset{\sim}{a}_{p-1}}) = S_{p-1}(F), \\
&S_p(F) = S(F^{\underset{\sim}{a}_p}) \quad \text{where} \quad |\underset{\sim}{a}_p| = 1,\ \underset{\sim}{a}_p \perp \underset{\sim}{a}_1,\ldots,\underset{\sim}{a}_{p-1}, \\
&A_p(F) = \underset{\sim}{a}_p\ ;
\end{aligned}
$$

or using the minimizing procedure:

$$S_p(F) = \min_{|\underset{\sim}{a}|=1} S(F^{\underset{\sim}{a}}) ,$$

$$A_p(F) = \underset{\sim}{a}_p \quad \text{where} \quad |\underset{\sim}{a}_p| = 1, \quad S(F^{\underset{\sim}{a}_p}) = S_p(F) ,$$

. . . . . .

$$S_2(F) = \min_{\substack{|\underset{\sim}{a}|=1 \\ \underset{\sim}{a} \perp \underset{\sim}{a}_p, \ldots, \underset{\sim}{a}_3}} \tag{1.1'}$$

$$A_2(F) = \underset{\sim}{a}_2 \quad \text{where} \quad |\underset{\sim}{a}_2| = 1, \ \underset{\sim}{a}_2 \perp \underset{\sim}{a}_p, \ldots, \underset{\sim}{a}_3, \ S(F^{\underset{\sim}{a}_2}) = S_2(F) ,$$

$$S_1(F) = S(F^{\underset{\sim}{a}_1}) , \quad \text{where} \ |\underset{\sim}{a}_1| = 1, \ \underset{\sim}{a}_1 \perp \underset{\sim}{a}_p, \ldots, \underset{\sim}{a}_2 ,$$

$$A_1(F) = \underset{\sim}{a}_1 .$$

Then the estimate for the covariance matrix, denoted by $\Sigma(F)$, is defined by

$$\Sigma(F) = \sum_{i=1}^{p} S_i(F)^2 A_i(F) A_i(F)^T . \tag{1.2}$$

Notice that here we can use "max" ("min") instead of "sup" ("inf") since we assume that $S(\cdot)$ is weakly continuous, thus for any F, $S(F^{\underset{\sim}{a}})$ is a continuous function of $\underset{\sim}{a}$ (see Lemma 4.4 (1) below), so there always exist $\underset{\sim}{a}_i$ $(1 \le i \le p)$ which reach the maximum (minimum) values in (1.1) [(1.1')] on the corresponding regions, respectively.

Obviously,

$$A_i(F) = \underset{\sim}{a}_i \Rightarrow A_i(F) = -\underset{\sim}{a}_i \quad (1 \le i \le p) . \tag{1.3}$$

So instead of $A_i(F)$ we often consider $A_i(F)A_i(F)^T$. Moreover, from (1.1) [(1.1')] one can see that $A_i(F)A_i(F)^T$, hence $\Sigma(F)$, may not be uniquely determined, so we account every possible version of $A_i(F)A_i(F)^T$ and $\Sigma(F)$.

Simulation results in [2] show that these robust PP-estimators compare favorably with the best robust estimators which have been proposed before. The robust PP-estimators have as good precision (measured by mean squared errors) as these best ones while achieving a higher (empirical) breakdown point.

This paper, as a companion of the simulation study [2], presents some theoretical results. Since weak continuity of $S(\cdot)$ is required, Section 2 shows that the M-estimates, in particular Huber's estimates, of scale are weakly continuous. Then in Section 3, we prove that the PP-estimator is orthogonally equivariant, and that at an elliptic underlying distribution it is also asymptotically affinely equivariant. Section 4 is devoted to the consistency at any distribution belonging to an elliptic probability density family. Qualitative and quantitative robustness are discussed in Section 5. It is shown that the robust PP-estimates are weakly continuous at elliptic densities and their breakdown point can be as high as 1/2. Finally, in Section 6, we make some relevant comments.

Since, for both the maximizing and minimizing procedures, the proofs of equivariance, consistency and the weak continuity are all the same, in Sections 3 and 4 and in the first part [illegible]on 5 we discuss only in the case of the maximizing procedure.

2. Further Study of M-Estimators for Scale

It is shown below that the robust PP-estimators will have good properties whenever the projection index, an estimator for scale, is weakly continuous and has a high breakdown point. Since M-estimators of scale are both simple and desirable, hence important, estimators, this section discusses their weak continuity and breakdown point.

Let $F(x)$ be one-dimensional distribution function, an M-estimator for scale of F, denoted by $S(F)$, is defined by an implicit equation

$$\int \chi\left(\frac{x}{S(F)}\right) dF(x) = 0 .$$

Usually $\chi(t)$ is even. $S(F)$, of course, possesses scale equivariance. Put

$$\lambda(s,F) = \int \chi\left(\frac{x}{s}\right) dF(x) \tag{2.1}$$

$$\tilde{F}(x) = \begin{cases} 0 & x \le 0 \\ F(x) - F(-x_{+0}) & x > 0 . \end{cases} \tag{2.2}$$

So $S(F)$ is a zero point of $\lambda(s,F)$. $\tilde{F}(x)$ is the distribution function of $|X|$ if $X \sim F(x)$.

LEMMA 2.1. Assume that $\chi(t)$ is even and increasing on $t > 0$ and $\lambda(s,F)$ exists for at least one $s \in R$. Then

(1) $\lambda(s,F) = \lambda(s,\tilde{F})$.

(2) For any fixed F, $\lambda(s,F)$ is decreasing in s; for any fixed s, $\lambda(s,F)$ is stochastically increasing in $\tilde{F}$.

(3) If $\chi(t)$ is also continuous and bounded then $\lambda(s,F)$ is continuous in s and weakly continuous in F for every s fixed.

Proof. (1) is very easy and (2) is straightforward from, for example, Lemma 2(i) in Lehmann's book (see [13], p. 74).

Now we prove (3). Denote Lévy metric by $d_L(\cdot,\cdot)$. For any small $\varepsilon > 0$, let

$$\mathscr{P}_\varepsilon(F_0) = \{F : d_L(F,F_0) < \varepsilon\} .$$

It is easy to check that

$$F \in \mathscr{P}_{\varepsilon/2}(F_0) \Rightarrow \tilde{F} \in \mathscr{P}_\varepsilon(\tilde{F}_0) .$$

Without losing generality, we assume that there exist x_1, x_2 $(0 < x_1$, $x_2 < \infty)$ satisfying

$$\tilde{F}_0(x_1) = \varepsilon , \quad \tilde{F}_0(x_2) = 1 - \varepsilon .$$

Let

$$F_1(x) = \begin{cases} 0 & x \le x_1 + \varepsilon \\ \tilde{F}_0(x-\varepsilon) - \varepsilon & x > x_1 + \varepsilon \\ 1 & x = \infty \end{cases} \tag{2.3}$$

$$F_2(x) = \begin{cases} 0 & x \le 0 \\ \tilde{F}_0(x+\varepsilon) + \varepsilon & 0 < x \le x_2 - \varepsilon \\ 1 & x > x_2 - \varepsilon . \end{cases}$$

Then for any $F \in \mathscr{P}_{\varepsilon/2}$, from (2) it follows that

$$\lambda(s,F) = \lambda(s,\tilde{F}) \le \lambda(s,F_1) = \int_{x_1+\varepsilon}^{\infty} \chi\left(\frac{t}{s}\right) d\tilde{F}_0(t-\varepsilon) + \varepsilon\chi(\infty)$$

$$= \int_{x_1}^{\infty} \chi\left(\frac{t+\varepsilon}{s}\right) d\tilde{F}_0(t) + \varepsilon\chi(\infty) = \int \chi\left(\frac{t+\varepsilon}{s}\right) d\tilde{F}_0(t)$$

$$- \int_0^{x_1} \chi\left(\frac{t+\varepsilon}{s}\right) d\tilde{F}_0(t) + \varepsilon\chi(\infty) \le \int \chi\left(\frac{t+\varepsilon}{s}\right) d\tilde{F}_0(t) + \varepsilon(\chi(\infty)-\chi(0)) .$$

$$\lambda(s,F) = \lambda(s,\tilde{F}) > \lambda(s,F_2) = \int_0^{x_2-\varepsilon} \chi\left(\frac{t}{s}\right) dF_0(t+\varepsilon) + (\varepsilon + F_0(\varepsilon)\chi(0))$$

$$= \int \chi\left(\frac{t-\varepsilon}{s}\right) dF_0(t) - \int_0^{\varepsilon} \chi\left(\frac{t-\varepsilon}{s}\right) dF_0(t) - \int_{x_2}^{\infty} \chi\left(\frac{t-\varepsilon}{s}\right) dF_0(t) + (\varepsilon + F_0(\varepsilon)\chi(0)$$

$$> \int \chi\left(\frac{t-\varepsilon}{s}\right) dF_0(t) - F_0(\varepsilon)\left(\chi\left(\frac{\varepsilon}{s}\right) - \chi(0)\right) - \varepsilon(\chi(\infty) - \chi(0)) .$$

Thus for any $0 < s < \infty$,

$$|\lambda(s,F)-\lambda(s,F_0)| \le \lambda(s,F_1) - \lambda(s,F_2) =$$

$$\le \int \left[\chi\left(\frac{t+\varepsilon}{s}\right) - \chi\left(\frac{t-\varepsilon}{s}\right)\right] dF_0(t) + 2\varepsilon(\chi(\infty)-\chi(0)) + F_0(\varepsilon)\left(\chi\left(\frac{\varepsilon}{s}\right) - \chi(0)\right)$$

$$\to 0 \qquad (\varepsilon \to 0) . \qquad \#$$

THEOREM 2.1. Assume that $\chi(t)$ is even, continuous, bounded and increasing on $t > 0$ and that $0 < S(F_0) < \infty$ is uniquely defined. Then $S(F)$ is weakly continuous at F_0.

Proof. Assume that $F_n \xrightarrow{w} F_0$. Put

$$s_i = S(F_i) \qquad i = 0, 1, 2, \ldots .$$

Then s_0 is the only zero point of $\lambda(s,F_0)$ and there exist s' and s'' $(s' < s_0 < s'')$ such that, from the monotonicity of $\lambda(\cdot,F_0)$ and uniqueness of $S(F_0)$,

$$\lambda(s'',F_0) < 0 < \lambda(s',F_0) .$$

From Lemma 2.1 it follows that for any $\varepsilon > 0$, there exist $t_0 = s' < t_1 < \ldots < t_m = s''$, and $N > 0$, such that

$$0 \le \lambda(t_{i-1},F_0) - \lambda(t_i,F_0) < \frac{\varepsilon}{2} \qquad i = 1, \ldots, m$$

$$|\lambda(t_i,F_n) - \lambda(t_i,F_0)| < \frac{\varepsilon}{2} \qquad n > N,\ i = 0, \ldots, m , \tag{2.4}$$

and also

$$\lambda(s'',F_n) < 0 < \lambda(s',F_n) \qquad n > N . \tag{2.5}$$

Hence, for any s, $s' \le s \le s''$, there exists an i, such that

$$t_{i-1} \le s \le t_i .$$

Then (2.4) gives that

$$\lambda(s,F_n) - \lambda(s,F_0) \le \lambda(t_{i-1},F_n) - \lambda(t_{i-1},F_0) + \lambda(t_{i-1},F_0) - \lambda(s,F_0)$$

$$\le \frac{\varepsilon}{2} + |\lambda(t_{i-1},F_n) - \lambda(t_{i-1},F_0)| \le \varepsilon$$

$$\lambda(s,F_n) - \lambda(s,F_0) \ge \lambda(t_i,F_n) - \lambda(t_i,F_0) + \lambda(t_i,F_0) - \lambda(s,F_0)$$

$$\ge -\frac{\varepsilon}{2} - \frac{\varepsilon}{2} = -\varepsilon .$$

That is, $\{\lambda(s,F) | s' \le s \le s''\}$ is weakly equicontinuous at F_0.

Now we claim that $s_n \to s_0$ $(n \to \infty)$. If not, because of (2.5), we have

$$s' \le s_n \le s'' .$$

Thus, there must be a subsequence of s_n, say $s_{n'}$, such that

$$s_{n'} \to \tilde{s} \ne s_0 .$$

On the other hand, because of the weak equicontinuity of $\lambda(s,F)$ at F_0,

$$\lambda(\tilde{s},F_0) = \lim_{n'\to\infty} \lambda(s_{n'},F_0) = \lim_{n'\to\infty} [\lambda(s_{n'},F_0) - \lambda(s_{n'},F_{n'})] = 0 .$$

This contradicts with the uniqueness of $S(F_0)$. Hence

$$S(F_n) = s_n \to S(F_0) \qquad (n \to \infty) . \qquad \blacksquare$$

Now we consider Huber's M-estimates of scale. That is the choice

$$\chi(t) = \begin{cases} t^2 - \beta & |t| < k \\ k^2 - \beta & |t| > k \end{cases} \tag{2.6}$$

with $k > 0$ and

$$\beta = \int_{|t|\leq k} t^2 \, d\Phi + k^2 \int_{|t|>k} d\Phi \tag{2.7}$$

where $\Phi(x)$ is the standard Normal distribution function. Obviously, $0 < \beta < k^2$ and $\chi(t)$ is even, bounded, continuous and increasing on $t > 0$. So it satisfies every condition in Lemma 2.1. Furthermore, we have

THEOREM 2.2. Assume that $F\{0\} = F(0_+) - F(0) < 1 - \frac{\beta}{k^2}$, then

(1) Huber's $S(F)$ is uniquely defined and $0 < S(F) < \infty$;

(2) $S(\cdot)$ is weakly continuous at F.

<u>Proof</u>. Put

$$x_0 = \sup\{x \mid x > 0, \ F(x_0) - F(0) = 0\} .$$

For any $s < \frac{x_0}{k}$,

$$\lambda(s,F) = \int_{|x|<ks} \left(\frac{x}{s}\right)^2 dF(x) + k^2 \int_{|x|>ks} dF(x) - \beta$$

$$= \int_{|x|>x_0} k^2 \, dF(x) - \beta = k^2(1-F\{0\}) - \beta > 0 . \tag{2.8}$$

If $\frac{x_0}{k} < s_1 < s_2$, then, notice that $F(x_0 < |x| < s_1 k) > 0$.

$$\lambda(s_2,F) = \int_{x_0 < |x| < s_1 k} \left(\frac{x}{s_2}\right)^2 dF(x)$$

$$+ \int_{s_1 k < |x| < s_2 k} \left(\frac{x}{s_2}\right)^2 dF(x) + k^2 \int_{|x| > s_2 k} dF(x) - \beta$$

$$< \int_{x_0 < |x| < s_1 k} \left(\frac{x}{s_1}\right)^2 dF(x) + \int_{|x| > s_1 k} k^2 \, dF(x) - \beta$$

$$= \lambda(s_1,F) . \tag{2.9}$$

It is clear that

$$\lim_{s\to\infty} \lambda(s,F) = -\beta < 0 . \tag{2.10}$$

Now (1) follows immediately from (2.8)-(2.10). And (2) is a necessary consequence of (1) and Theorem 2.1. $\blacksquare$

The breakdown point for M-estimates of scale has been worked out in the usual sense (see [10], p. 110). But we need to add something else to serve our purpose. For completeness, we discuss it from the beginning.

Suppose that $\chi(t)$ is even and increasing on $t > 0$. Then Lemma 2.1 holds. Put

$$\begin{aligned} \|\chi\| &= \chi(\infty) - \chi(0), \\ \lambda(\infty,F) &= \lim_{s\to\infty} \lambda(s,F) . \end{aligned} \tag{2.11}$$

Consider the gross error model

$$F = (1-\epsilon)F_0 + \epsilon H .$$

Since "$S(F) < \infty$ for any H" implies that $\lambda(\infty, F) \leq 0$ for any H, we have

$$\lambda(\infty,\ (1-\varepsilon)F_0+\varepsilon\delta_\infty) = (1-\varepsilon)\chi(0) + \varepsilon\chi(\infty) \leq 0 ,$$

where δ_x is a probability measure putting a pointmass at x. This is equivalent to

$$\varepsilon \leq -\frac{\chi(0)}{\|\chi\|} = \varepsilon_1 . \tag{2.12}$$

It is easy to show that if $\|\chi\| < \infty$, then

$$\varepsilon < -\frac{\chi(0)}{\|\chi\|} \Rightarrow S(F) < \infty \quad \text{for any } H.$$

Conversely, if $\varepsilon > \frac{-\chi(0)}{\|\chi\|}$, then

$$H = \delta_\infty \Rightarrow \lambda(\infty, F) > 0 \Rightarrow S(F) = \infty .$$

Sometime we hope that $S(F) > 0$ when F_0 is not degenerate (into 0). In this case, if $S(F) > 0$ for any H, then

$$\lambda(0_+,\ (1-\varepsilon)F_0+\varepsilon\delta_0) - (1-\varepsilon)\{F_0\{0\}\chi(0)+(1-F_0\{0\})\chi(\infty)\} + \varepsilon\chi(0) \geq 0 ,$$

i.e.,

$$\varepsilon < \frac{\chi(\infty)}{\|\chi\|} + \frac{F_0\{0\}\chi(0)}{(1-F_0\{0\})\|\chi\|} = \varepsilon_2 . \tag{2.13}$$

Also

$$\varepsilon < \varepsilon_2 \Rightarrow \lambda(0_+,\ (1-\varepsilon)F_0+\varepsilon\delta_0) > 0 \Rightarrow S(F) > 0 \quad \text{for any } H.$$

On the other hand,

$$\varepsilon > \varepsilon_2 \Rightarrow \lambda(0_+,\ (1-\varepsilon)F_0+\varepsilon\delta_0) < 0 \Rightarrow S((1-\varepsilon)F_0+\varepsilon\delta_0) = 0 .$$

As we will see in the end of Section 5, it is important that $S(F)$ does not misbehave as a projection index for principal component estimation. That

means, if F_0 is degenerate (into 0) then $S(F) = 0$ for any H. This implies that

$$\lambda(0_+,\ (1-\varepsilon)F_0+\varepsilon\delta_\infty) = (1-\varepsilon)\chi(0) + \varepsilon\chi(\infty) \leq 0 , \tag{2.14}$$

that is,

$$\varepsilon \leq \frac{-\chi(0)}{\|\chi\|} = \varepsilon_3 \ (= \varepsilon_1) . \tag{2.15}$$

Also

$$\varepsilon < \varepsilon_3 \Rightarrow \lambda(0_+, F) < 0 \quad \text{for any } H \Rightarrow S(F) = 0 \quad \text{for any } H.$$

Conversely, if

$$\varepsilon > \varepsilon_3$$

then

$$\lambda(0_+,\ (1-\varepsilon)F_0+\varepsilon\delta_\infty) = \lambda(0_+,\ (1-\varepsilon)\delta_0+\varepsilon\delta_\infty) > 0 \Rightarrow S((1-\varepsilon)\delta_0+\varepsilon\delta_\infty) > 0 .$$

As Huber pointed out (see [10], p. 110) that we can usually disregard the second contingency, so we conclude that for the ε-contamination model, the breakdown point of M-estimates for scale is

$$\varepsilon^* = \frac{-\chi(0)}{\|\chi\|} .$$

In the case of Huber's choice,

$$\varepsilon^* = \frac{\beta}{k^2} .$$

Now we see obviously that for any even $\chi(t)$, the breakdown point of an M-estimator for scale is as high as 1/2.

3. Equivariance

The study of robust estimators of covariance poses a problem: How to provide a covariance estimator which behaves under coordinate changes the same way as the classical estimator does. It is important to establish equivariance properties of a new estimator in order to show that it is, truly, an estimator of dispersion. This section shows that the robust PP-estimators do have equivariance properties.

As it is mentioned in Section 1 that this paper considers the pure dispersion problem, i.e., assume that location is known and fixed at $\mathbf{0}$. Also, we assume throughout that $S(\cdot)$ is weakly continuous.

Recalling that $A_i(F)$ and Σ are not uniquely determined, we use them to denote any version of those quantities.

THEOREM 3.1. Robust PP-estimates $S_i(F)$, $A_i(F)$ $(i = 1, 2, \ldots, p)$ and $\Sigma(F)$ are all orthogonal equivariant. I.e., let P be any orthogonal matrix, $\mathbf{X} \sim F_1(\mathbf{x})$ and $\mathbf{Y} = P\mathbf{X} \sim F_2(\mathbf{x})$; then

$$S_i(F_2) = S_i(F_1)$$

$$A_i(F_2) = PA_i(F_1)$$

$$\Sigma(F_2) = P\Sigma(F_1)P^T \qquad i = 1, 2, \ldots, p .$$

Proof. Since

$$S_1(F_1) = \max_{|\boldsymbol{\alpha}|=1} S(\mathcal{L}(\boldsymbol{\alpha}^T\mathbf{X})) = \max_{|\boldsymbol{\alpha}|=1} S\{\mathcal{L}\{(P\boldsymbol{\alpha})^T\mathbf{Y}\}\}$$

$$= \max_{|\boldsymbol{\beta}|=1} S(\mathcal{L}(\boldsymbol{\beta}^T\mathbf{Y})) = S_1(F_2) . \tag{3.1}$$

Thus

$$S_1(F_2) = S_1(F_1) = S[\mathcal{L}(A_1(F_1)^T\mathbf{X})] = S[\mathcal{L}((PA_1(F_1))^T\mathbf{Y})] ,$$

i.e.,

$$A_1(F_2) = PA_1(F_1) . \tag{3.2}$$

From (3.1) and (3.2), it follows that

$$S_2(F_1) = \max_{\substack{|\boldsymbol{\alpha}| = 1 \\ \boldsymbol{\alpha} \perp A_1(F_1)}} S(\mathcal{L}(\boldsymbol{\alpha}^T\mathbf{X})) = \max_{\substack{|\boldsymbol{\alpha}| = 1 \\ \boldsymbol{\alpha} \perp P^TA_1(F_2)}} S[\mathcal{L}((P\boldsymbol{\alpha})^T\mathbf{Y})]$$

$$= \max_{\substack{P\boldsymbol{\alpha} \perp A_1(F_2) \\ |P\boldsymbol{\alpha}| = 1}} S[\mathcal{L}((P\boldsymbol{\alpha})^T\mathbf{Y})] = \max_{\substack{|\boldsymbol{\beta}| = 1 \\ \boldsymbol{\beta} \perp A_1(F_2)}} S(\mathcal{L}(\boldsymbol{\beta}^T\mathbf{Y})) = S_2(F_2) ,$$

$$S_2(F_2) = S_2(F_1) = S[\mathcal{L}(A_2(F_1)^T\mathbf{X})] = S[\mathcal{L}((PA_2(F_1))^T\mathbf{Y})] ,$$

$$\Rightarrow A_2(F_2) = PA_2(F_1) .$$

And so on, we have

$$S_i(F_2) = S_i(F_1)$$

$$A_i(F_2) = PA_i(F_1) \qquad i = 1, 2, \ldots, p .$$

Therefore,

$$\Sigma(F_2) = \sum_i S_i(F_2)^2 A_i(F_2)A_i(F_2)^T$$

$$= \sum_i S_i(F_1)^2 PA_i(F_1)A_i(F_1)^T P^T$$

$$= P\,\Sigma(F_1)P^T . \qquad \blacksquare$$

Notice that a random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ and the associated empirical distribution $F_n(\mathbf{x}) = \frac{1}{n}\sum_{1}^{n} \delta_{\mathbf{x}_i}(\mathbf{x})$ change, under a coordinate transformation

P, into $PX_1, \ldots, PX_n$ and

$$G_n(x) = \frac{1}{n}\sum_1^n \delta_{PX_i}(x) = \frac{1}{n}\sum_1^n \delta_{X_i}(P^{-1}x) = F_n(P^{-1}x) ,$$

respectively, and that

$$X \sim F_n(x) \Rightarrow PX \sim F_n(P^{-1}x) .$$

So from Theorem 3.1, it follows that as an estimator from sample

$$S_i(x_1,\ldots,x_n) = S_i(F_n) \qquad i = 1, \ldots, p ,$$

$$A_i(x_1,\ldots,x_n) = A_i(F_n) \qquad i = 1, \ldots, p ,$$

$$\Sigma(x_1,\ldots,x_n) = \Sigma(F_n)$$

are indeed orthogonal equivariant, just like classical estimators.

Suppose that $F(x,V)$ belongs to a p-dimension elliptic probability density family, i.e., $F(x,V)$ has a density $f(x,V)$

$$f(x,V) = (\det V^{-1}) f_0(V^{-1}x) , \tag{3.3}$$

where V is a nonsingular $p \times p$ matrix and $f_0(x)$ is spherically symmetric (and nondegenerate, of course), i.e.,

$$f_0(x) = f_0(|x|) . \tag{3.4}$$

To prove the affine equivariance within an elliptic density family, we need

LEMMA 3.1. Assume that p-dimensional random vector X has a spherically symmetric probability density $f_0(x) = f_0(x_1,\ldots,x_p) = f_0(|x|)$. Let

$$g(x) = \int f_0(x,x_2,\ldots,x_p)\, dx_2 dx_3 \ldots dx_p \tag{3.5}$$

be the marginal density. Then $a^T X \sim g(x)$ for any unit vector a.

Proof. For any unit a fixed, let P be an orthogonal matrix such that

$$P = \begin{pmatrix} a^T \\ \vdots \end{pmatrix} ,$$

and put $Y = PX = (Y_1,\ldots,Y_p)^T$.

Since Y has a probability density

$$f_Y(y) = (\det P^{-1}) f_0(P^{-1}y) = f_0(|P^{-1}y|) = f_0(y) ,$$

it is immediate that

$$a^T X = Y_1 \sim g(x) . \qquad \blacksquare$$

Let $F_0(x)$ be probability distribution function of $f_0(x)$ and let $G(x)$ be $g(x)$'s [see (3.5)]. From Lemma 3.1, it is straightforward that

$$S_i(F_0) = \min_{\substack{|a| = 1 \\ a \perp A_1(F_0),\ldots,A_{i-1}(F_0)}} S(\mathcal{L}(a^T X)) = S(G) \qquad i = 1, 2, \ldots, p ;$$

and any set of orthonormal vectors $a_1, \ldots, a_p$ can be $A_1(F_0), \ldots, A_p(F_0)$. Without losing generality, we assume that

$$S(G) = 1 . \tag{3.6}$$

Thus,

$$\Sigma(F_0) = \sum_i S_i(F_0)^2 A_i(F_0) A_i(F_0)^T = S(G) I = I . \tag{3.7}$$

Also, Lemma 3.1, together with the scale equivariance of $S(\cdot)$, yields that for any $a \in R^p$, $a \neq 0$,

$$S(\mathcal{L}(a^T X)) = S\left[\mathcal{L}\left(|a| \frac{a^T X}{|a|}\right)\right] = |a| . \tag{3.8}$$

Hence, for any member of the elliptic family $F_V(\underset{\sim}{x}) = F(\underset{\sim}{x}, V)$ [see (3.3) and (3.4)], put

$$\begin{aligned} \underset{\sim}{Y} &= V\underset{\sim}{X}\,, \\ \Sigma &= VV^T\,, \end{aligned} \qquad (3.9)$$

Σ is usually called the pseudo-covariance matrix of $\underset{\sim}{X}$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$ be the eigenvalues of Σ and $\underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_p$ be any set of eigenvectors corresponding to $\lambda_1, \ldots, \lambda_p$, respectively. (3.8) gives us that for any $\underset{\sim}{a} \in R^p$,

$$S(F_V^{\underset{\sim}{a}}) = S(\mathscr{F}(\underset{\sim}{a}^T\underset{\sim}{Y})) = S[\mathscr{F}((V^T\underset{\sim}{a})^T\underset{\sim}{X})] = |V^T\underset{\sim}{a}|\,.$$

Therefore, simple algebra gives

$$\begin{aligned} S_1(F_V) &= \max_{|\underset{\sim}{a}|=1} S(F_V^{\underset{\sim}{a}}) = \max_{|\underset{\sim}{a}|=1} |V^T\underset{\sim}{a}| = |V^T\underset{\sim}{\alpha}_1| = \sqrt{\lambda_1}\,, \\ A_1(F_V) &= \underset{\sim}{\alpha}_1\,, \\ S_2(F_V) &= \max_{\substack{|\underset{\sim}{a}|=1 \\ \underset{\sim}{a} \perp \underset{\sim}{\alpha}_1}} S(F_V^{\underset{\sim}{a}}) = |V^T\underset{\sim}{\alpha}_2| = \sqrt{\lambda_2}\,, \\ A_2(F_V) &= \underset{\sim}{\alpha}_2\,, \\ &\cdots\cdots\cdots \\ S_p(F_V) &= \sqrt{\lambda_p}\,, \\ A_p(F_V) &= \underset{\sim}{\alpha}_p\,. \end{aligned} \qquad (3.10)$$

Consequently,

$$\mathfrak{S}(F_V) = \sum_i S_i^2(F_V)A_i(F_V)A_i(F_V)^T = \sum_i \lambda_i \underset{\sim}{\alpha}_i \underset{\sim}{\alpha}_i^T = \Sigma\,.$$

So far, we have proved

THEOREM 3.2. Within any elliptic probability density family, robust PP-estimator for covariance matrix $\mathfrak{S}(F)$, as a functional, is affinely equivariant. In detail, assume that $\underset{\sim}{X} \sim F(\underset{\sim}{x})$ belongs to an elliptic density family and that V is any nondegenerate $p \times p$ matrix. Denote the distribution of $V\underset{\sim}{X}$ by $G(\underset{\sim}{x})$. Then

$$\mathfrak{S}(G) = V\mathfrak{S}(F)V^T.$$

Under an affine transformation V, an empirical distribution $F_n(\underset{\sim}{x}) = \frac{1}{n}\sum_1^n \delta_{\underset{\sim}{x}_i}(\underset{\sim}{x})$ transforms into $G_n(\underset{\sim}{x}) = \frac{1}{n}\sum_1^n \delta_{V\underset{\sim}{x}_i}(\underset{\sim}{x}) = \frac{1}{n}F_n(V^{-1}\underset{\sim}{x})$. Since $F_n(\underset{\sim}{x})$ is generally not generated by a spherical distribution, $F_n(\underset{\sim}{x})$ and $G_n(\underset{\sim}{x})$ do not belong to any elliptic distribution family. Therefore, as an estimate from sample, $\mathfrak{S}(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n) = \mathfrak{S}(F_n)$ is not affinely equivariant. However, if the underlying distribution of F_n is a member of an elliptic family, then $\mathfrak{S}(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n) = \mathfrak{S}(F_n)$ is asymptotically affinely equivariant. Actually, suppose $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$ come from an elliptic distribution $F(\underset{\sim}{x}, V_1)$, according to Theorem 4.4,

$$\mathfrak{S}(\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n) = \mathfrak{S}(F_n) \to \Sigma_1 = \mathfrak{S}(F(\underset{\sim}{x}, V_1)) \qquad \text{P. \& a. s.}$$

$$\mathfrak{S}(V\underset{\sim}{x}_1,\ldots,V\underset{\sim}{x}_n) = \mathfrak{S}(G_n) \to \Sigma_2 = \mathfrak{S}(F(\underset{\sim}{x}, V_2)) \qquad \text{P. \& a. s.}\,;$$

then, from Theorem 3.2, it follows that

$$\Sigma_2 = V\Sigma_1V^T\,.$$

4. Consistency at Elliptic Probability Density Family

When the data come from an elliptic probability distribution, for which pseudocovariance has an intuitive interpretation — the shape of the underlying ellipse, it is possible to show that the robust PP-estimators give consistent estimates. The idea of the proof is simple: it combines the continuity property of the projection index $S(\cdot)$ (discussed in Section 3 for M-estimates) with a compactness argument. But putting this into operation in high dimension and for any underlying pseudo-covariance is quite complicated. To cope with this problem, we introduce some lemmas first, then start with some special underlying pseudo-covariances to achieve general consistency results.

LEMMA 4.1. Assume that

(1) Ω is a compact set in R^p, $\mathscr{S}_n(\underset{\sim}{\alpha})$ $(n = 0, 1, 2, \ldots)$ are continuous on Ω.

(2) $\lim\limits_{n\to\infty} \mathscr{S}_n(\underset{\sim}{\alpha}) = \mathscr{S}_0(\underset{\sim}{\alpha})$ uniformly in $\underset{\sim}{\alpha}$ on Ω.

(3) For $\Omega_n \subset \Omega$ $(n = 0, 1, 2, \ldots)$, there exist $\underset{\sim}{\alpha}_n$ such that

$$\mathscr{S}_n(\underset{\sim}{\alpha}_n) = \max_{\Omega_n} \mathscr{S}_n(\underset{\sim}{\alpha}) \qquad n = 0, 1, 2, \ldots .$$

(4) There also exist $p \times p$ orthogonal matrices P_n such that

$$\Omega_n = P_n\Omega_0 = \{P_n\underset{\sim}{\alpha} \mid \underset{\sim}{\alpha} \in \Omega_0\} \qquad n = 1, 2, \ldots$$

$$P_n \to I \qquad (n \to \infty)$$

(here we use $\|A\| = \mathrm{trace}(AA^T)$ as the norm of matrix A).

Then

$$\mathscr{S}_n(\underset{\sim}{\alpha}_n) \to \mathscr{S}_0(\underset{\sim}{\alpha}_0) \qquad (n \to \infty) .$$

Proof. Notice that

$$P_n\underset{\sim}{\alpha}_0 \in \Omega_n , \quad P_n^{-1}\underset{\sim}{\alpha}_n \in \Omega_0 , \qquad n = 1, 2, \ldots .$$

From assumption (3) we have

$$\mathscr{S}_n(P_n\underset{\sim}{\alpha}_0) - \mathscr{S}_0(\underset{\sim}{\alpha}_0) < \mathscr{S}_n(\underset{\sim}{\alpha}_n) - \mathscr{S}_0(\underset{\sim}{\alpha}_0) < \mathscr{S}_n(\underset{\sim}{\alpha}_n) - \mathscr{S}_0(P_n^{-1}\underset{\sim}{\alpha}_0) .$$

Now we need only show that the two extreme sides of this inequality converge to zero when $n \to \infty$. Actually, by assumption (4), (1) and (2), it follows that

$$P_n\underset{\sim}{\alpha}_0 \to \underset{\sim}{\alpha}_0 \qquad (n \to \infty)$$

$$|\mathscr{S}_n(P_n\underset{\sim}{\alpha}_0) - \mathscr{S}_0(\underset{\sim}{\alpha}_0)| < |\mathscr{S}_n(P_n\underset{\sim}{\alpha}_0) - \mathscr{S}_0(P_n\underset{\sim}{\alpha}_0)| + |\mathscr{S}_0(P_n\underset{\sim}{\alpha}_0) - \mathscr{S}_0(\underset{\sim}{\alpha}_0)| \to 0 \quad (n \to \infty) .$$

Similarly,

$$|\mathscr{S}_n(\underset{\sim}{\alpha}_n) - \mathscr{S}_0(P_n^{-1}\underset{\sim}{\alpha}_0)| \to 0 \quad (n \to \infty) . \qquad \blacksquare$$

LEMMA 4.2. Under the conditions of Lemma 4.1, if $\underset{\sim}{\alpha}_0$ is the unique maximum of $\mathscr{S}_0(\underset{\sim}{\alpha})$ on $\bar{\Omega}_0$, the closure of Ω_0, then

$$\underset{\sim}{\alpha}_n \to \underset{\sim}{\alpha}_0 \qquad n \to \infty .$$

Proof. Put

$$\underset{\sim}{\beta}_n = P_n^{-1}\underset{\sim}{\alpha}_n ,$$

by the assumption (4) in Lemma 4.1, we know that

$$\beta_n \in \Omega_0$$

$$\lim_{n\to\infty} (\beta_n - \alpha_n) = 0 .$$

Assume that

$$\alpha_n \not\to \alpha_0 \qquad n \to \infty .$$

Since $\alpha_n \in \Omega_n \subset \Omega$ and Ω is compact, there must be a subsequence of α_n, $\alpha_{n'}$, such that $\lim_{n'\to\infty} \alpha_{n'} = \alpha$. Thus $\lim_{n'\to\infty} \beta_{n'} = \alpha$ and $\alpha \in \bar{\Omega}_0$. Then assumptions (1)-(3) give

$$\mathscr{S}_0(\alpha) = \lim_{n'\to\infty} \mathscr{S}_0(\alpha_{n'}) = \lim_{n'\to\infty} \mathscr{S}_{n'}(\alpha_{n'}) \geq \lim_{n'\to\infty} \mathscr{S}_{n'}(\alpha_0) = \mathscr{S}_0(\alpha_0) .$$

This contradicts that α_0 is the only maximum of $\mathscr{S}_0(\cdot)$ on $\bar{\Omega}_0$. Hence

$$\alpha_n \to \alpha_0 \qquad (n \to \infty) . \qquad \blacksquare$$

LEMMA 4.3.

(1) Assume α, $\alpha_1 \in R^p$, $|\alpha| = |\alpha_1| = 1$, $\alpha_1 \neq \alpha$, then there exists a rotation matrix P such that

$$\begin{aligned} &\alpha_1 = P\alpha \\ &\beta = P\beta \quad \text{for any } \beta \perp \alpha, \alpha_1 \\ &|P\beta - \beta| \leq |\alpha_1 - \alpha| \quad \forall \beta . \end{aligned} \tag{4.1}$$

(2) Assume that α_n, $\alpha \in R^p$, $|\alpha_n| = |\alpha| = 1$, and that P_n is a rotation matrix as in (1) such that

$$\alpha_n = P_n\alpha \qquad n = 1, 2, \ldots .$$

Then $P_n \to I$ $(n \to \infty) \Longleftrightarrow \alpha_n \to \alpha$ $(n \to \infty)$.

Proof.

(1) Without losing generality, we assume that α, α_1 are on the same coordinate plane formed by the first two axes. Let θ be the angle between α and α_1, we define a matrix P by

$$P = \begin{pmatrix} \cos\theta & -\sin\theta & \\ \sin\theta & \cos\theta & 0 \\ & 0 & I_{p-2} \end{pmatrix} . \tag{4.2}$$

It is easy to check that P is the rotation matrix we want, i.e., it satisfies (4.1).

(2) Let θ_n be the angle between α and α_n $(n = 1, 2, \ldots)$. From (1) we know that

$$P_n = \begin{pmatrix} \cos\theta_n & -\sin\theta_n & \\ \sin\theta_n & \cos\theta_n & 0 \\ & 0 & I_{p-2} \end{pmatrix} .$$

Obviously,

$$\underset{(n\to\infty)}{\alpha_n \to \alpha} \Longleftrightarrow \underset{(n\to\infty)}{\theta_n \to 0} \Longleftrightarrow \underset{(n\to\infty)}{P_n \to I} . \tag{4.3}$$

LEMMA 4.4. Assume that $X \sim F_0(x)$ is a P-dimension random vector and $F_n(x)$ $(n = 1, 2, \ldots)$ are the empirical distribution of F_0, and that M-estimate $S(\cdot)$ for scale is weakly continuous. Then

(1) for every n $(n = 0, 1, 2, \ldots)$ fixed, F_n^{α} is weakly continuous in α, $S(F_n^{\alpha})$ is continuous in α;

(2) $\mathscr{F} = \{F_n^{\alpha} | \alpha \in R^p,\ |\alpha| = 1,\ n = 0,1,2,\ldots\}$ is a.s. weakly compact;

(3) if $|\alpha_n| = |\beta_n| = 1$, $(\alpha_n - \beta_n) \rightarrow 0\ (n \rightarrow \infty)$, then

$$S(F_0^{\alpha_n}) - S(F_0^{\beta_n}) \rightarrow 0 \quad (n \rightarrow \infty)$$

$$S(F_n^{\alpha_n}) - S(F_0^{\beta_n}) \rightarrow 0 \quad (n \rightarrow \infty) \qquad \text{P. \& a.s.}$$

$$S(F_n^{\alpha_n}) - S(F_n^{\beta_n}) \rightarrow 0 \quad (n \rightarrow \infty) \qquad \text{P. \& a.s.} \quad .$$

<u>Proof</u>.

(1) For any n (n = 0, 1, 2, ...) fixed, assume random vector $Y \sim F_n$; thus, according to our convention, $\alpha^T Y \sim F_n^{\alpha}$. If $\alpha_t \rightarrow \alpha\ (t \rightarrow \infty)$, then

$$\alpha_t^T Y \rightarrow \alpha^T Y \quad (t \rightarrow \infty) \quad \text{everywhere.}$$

Thus

$$F_n^{\alpha_t} \rightarrow F_n^{\alpha}\ (W) \quad (t \rightarrow \infty) \ .$$

By the weak continuity of $S(\cdot)$, we have

$$S(F_n^{\alpha_t}) \rightarrow S(F_n^{\alpha}) \quad (t \rightarrow \infty) \ .$$

(2) We have to show that any sequence $\{G_t | t=1,2,\ldots\} \subset \mathscr{F}$ has a subsequence which a.s. converges according to weak topology. Suppose that

$$G_t = F_{n_t}^{\alpha_t} \in , \qquad t = 1, 2, \ldots \ .$$

If there are only finite different n_i in $\{n_t | t=1,2,\ldots\}$, denote these n_i which are the same by m, then choose a subsequence of α_t, namely $\alpha_{t'}$, such

that $\alpha_{t'} \rightarrow \alpha\ (t' \rightarrow \infty)$ and $n_{t'} = m$. From (1) we obtain that

$$G_{t'} = F_m^{\alpha_{t'}} \rightarrow F_m^{\alpha} \ \text{a.s.} \quad (t' \rightarrow \infty) \ .$$

Now assume that $\{n_t | t=1,2,\ldots\}$ has infinitely many different elements. We can choose a subsequence of α_t, say $\alpha_{t'}$, such that

$$\alpha_{t'} \rightarrow \alpha \ \text{and}\ n_{t'} \rightarrow \infty \quad (t' \rightarrow \infty) \ .$$

Then, let d_* is a metric which measures weak topology, we know

$$d_*(G_{t'}, F_0^{\alpha}) \leq d_*(F_{n_{t'}}^{\alpha_{t'}}, F_0^{\alpha_{t'}}) + d_*(F_0^{\alpha_{t'}}, F_0^{\alpha}) \ . \tag{4.4}$$

Because $F_0^{\alpha_{t'}} \rightarrow F_0^{\alpha}\ (W)\ (t' \rightarrow \infty)$, the second term of the right-hand side of (4.4) vanishes when $t' \rightarrow \infty$. Denote the Kolmogorov metric by $d_k(\cdot,\cdot)$. From the Glivenko-Cantelli theorem, i.e.,

$$P[d_k(H_n, H) \rightarrow 0,\ n \rightarrow \infty] = 1$$

uniformly in H (H_n are empirical distributions of H), it follows that

$$d_*(F_{n_{t'}}^{\alpha_{t'}}, F_0^{\alpha_{t'}}) \leq d_k(F_{n_{t'}}^{\alpha_{t'}}, F_0^{\alpha_{t'}}) \rightarrow 0 \ \text{a.s.} \quad (t' \rightarrow \infty) \ . \tag{4.5}$$

Hence, the first term of the right-hand side of (4.4) also converges to zero when $t' \rightarrow \infty$. Thus

$$G_{t'} \rightarrow F_0^{\alpha}\ (W) \ \text{a.s.} \quad (t' \rightarrow \infty) \ .$$

(3) As a result of (1), $S(F_0^{\alpha})$ is uniformly continuous on $\{\alpha | |\alpha|=1,\ \alpha \in R^p\}$. It follows immediately that

$$|\alpha_n - \beta_n| \rightarrow 0\ (n \rightarrow \infty) \Rightarrow S(F_0^{\alpha_n}) - S(F_0^{\beta_n}) \rightarrow 0 \quad (n \rightarrow \infty) \ .$$

To complete (3), we have only to show that

$$S(F_n^{\underline{\alpha}_n}) - S(F_0^{\underline{\beta}_n}) \to 0 \ (n \to \infty) , \quad P. \ \& \ a.s. \quad .$$

But

$$S(F_n^{\underline{\alpha}_n}) - S(F_0^{\underline{\beta}_n}) = [S(F_n^{\underline{\alpha}_n}) - S(F_0^{\underline{\alpha}_n})] + [S(F_0^{\underline{\alpha}_n}) - S(F_0^{\underline{\beta}_n})] .$$

We know already that the second term of the right-hand side vanishes when $n \to \infty$; as for the first term, since $S(\cdot)$ is weakly continuous, $\mathcal{W}$ is a.s. weakly compact, so $S(\cdot)$ is uniformly continuous a.s. on $\mathcal{W}$; for verifying of

$$S(F_n^{\underline{\alpha}_n}) - S(F_0^{\underline{\alpha}_n}) \to 0 \quad n \to \infty \quad P. \ \& \ a.s.$$

we need only

$$d_*(F_n^{\underline{\alpha}_n}, F_0^{\underline{\alpha}_n}) \to 0 \quad n \to \infty \quad P. \ \& \ a.s. \ .$$

This is straightforward from (4.5). ∎

COROLLARY. Put

$$\Omega = \{\underline{\alpha} | \underline{\alpha} \in R^p, \ |\underline{\alpha}| = 1\} . \tag{4.6}$$

$S(F_n^{\underline{\alpha}})$ weakly converges (P. & a.s.) to $S(F_0^{\underline{\alpha}})$ $(n \to \infty)$ uniformly in $\underline{\alpha}$ on Ω.

Roughly speaking, Lemmas 4.1 and 4.2 say that under certain conditions, the maximum values and the maximum points of a convergent sequence converge to the maximum value and the maximum point of the limiting function. This would be the basic idea of the consistency of our robust PP-estimates, since $S_i(F_n)$, $A_i(F_n)$, $S_i(F)$ and $A_i(F)$ are maximum values and maximum points.

Lemmas 4.3 and 4.4 will essentially provide the assumptions in Lemmas 4.1 and 4.2 for $S(F_n^{\underline{a}})$ and $S(F^{\underline{a}})$. Especially when the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$ are all different, it is easy to make $A_i(F)$ $(1 \le i \le p)$ unique on certain regions; thus Lemma 4.2 holds for $S(F_n^{\underline{a}})$ and $S(F^{\underline{a}})$. Then the consistency of $S_i(F_n)$ and $A_i(F_n)$, hence $\hat{\Sigma}(F_n)$, are almost straightforward.

Now, let F_n $(n = 1,2,\ldots)$ be the empirical distribution of $F = F(\underline{\mu},V)$, and recall the notations $F(\underline{\mu},V)$, Σ, λ_i, and $\underline{\alpha}_i$ introduced in Section 3 (see somewhere around (3.3) and (3.9)). To avoid the unnecessary multiple solutions of $A_i(\cdot)$ caused by (1.3), we can reduce the regions over which the maximum values $S_i(\cdot)$ are obtained. Let G be any P-dimension distribution; put

$$s_i = S_i(G) , \quad a_i = A_i(G) ,$$

$$D_1 = \{\underline{a} | \underline{a} \in R^p, \ |\underline{a}| = 1\} ,$$

$$D_i = \{\underline{a} | \underline{a} \in R^p, \ |\underline{a}| = 1, \ \underline{a} \perp \underline{a}_1, \ldots, \underline{a}_{i-1}\} \quad i = 2, \ldots, p .$$

Let W_i be any half of D_i satisfying

$$\begin{aligned} &\underline{a} \in W_i \Rightarrow -\underline{a} \notin W_i , \\ &W_i \cup (-W_i) = D_i \end{aligned} \tag{4.7}$$

where $(-W_i) = \{-\underline{a} | \underline{a} \in W_i\}$. For simplicity, we call such a half of a hypersphere a SKEW HALF. From (1.3), it follows that W_i contains at least one maximum point of $S(G^{\underline{a}})$ over D_i, and

$$S_i(G) = \max_{W_i} S(G^{\underline{a}}) .$$

Put

$$s_{in} = S_i(F_n)$$

and let $\underset{\sim}{\alpha}_{in}$ be any value of $A_i(F_n)$ in a certain skew half (we will specify it later) of the (P-i+1)-dimension hypersphere, and Σ_n be the associated value of $\Sigma(F_n)$. We are going to prove the consistency step by step.

THEOREM 4.1. Assume that $\lambda_1 > \lambda_2 > \ldots > \lambda_p > 0$, then

$$s_{in} \to \sqrt{\lambda_i}$$

$$\underset{\sim}{\alpha}_{in}\underset{\sim}{\alpha}_{in}^T \to \underset{\sim}{\alpha}_i\underset{\sim}{\alpha}_i^T \qquad (n \to \infty), \qquad P. \ F \text{ a.e.}\ .$$

$$\Sigma_n \to \Sigma$$

<u>Proof</u>. First of all, we adjust our problem to Lemma 4.2, as we have mentioned before. Since the eigenvalues $\lambda_1, \ldots, \lambda_p$ are all different, $A_i(F)$ has only two possible values $\pm\underset{\sim}{\alpha}_i$, and any skew half of hypersphere $B_i = \{\underset{\sim}{\alpha} | \underset{\sim}{\alpha} \in R^P, |\underset{\sim}{\alpha}|=1, \underset{\sim}{\alpha} \perp \underset{\sim}{\alpha}_1, \ldots, \underset{\sim}{\alpha}_{i-1}\}$, say Q_i, contains only one of the two maximum points $\pm\underset{\sim}{\alpha}_i$. We choose Q_i containing $\underset{\sim}{\alpha}_i$. Lemma 4.2 also requires that the maximum is unique over the closure of that region, so we need to make sure that

$$-\underset{\sim}{\alpha}_i \notin \bar{Q}_i\ .$$

Specifically, we choose Q_i such that $\underset{\sim}{\alpha}_i$ is its "center." In detail, let

$$\begin{aligned}
&Q_p = \{\underset{\sim}{\alpha}_p\}\ ,\\
&Q_{p-1} = \{\underset{\sim}{\alpha} \,|\, |\underset{\sim}{\alpha}|=1,\ \underset{\sim}{\alpha}^T\underset{\sim}{\alpha}_i=0,\ i=1,2,\ldots,p-2;\ \underset{\sim}{\alpha}^T\underset{\sim}{\alpha}_{p-1}>0\} \cup Q_p\ ,\\
&Q_{p-2} = \{\underset{\sim}{\alpha} \,|\, |\underset{\sim}{\alpha}|=1,\ \underset{\sim}{\alpha}^T\underset{\sim}{\alpha}_i=0,\ i=1,2,\ldots,p-3;\ \underset{\sim}{\alpha}^T\underset{\sim}{\alpha}_{p-2}>0\} \cup Q_{p-1}\ ,\\
&\cdots\cdots\cdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4.8)\\
&Q_2 = \{\underset{\sim}{\alpha} \,|\, |\underset{\sim}{\alpha}|=1,\ \underset{\sim}{\alpha}^T\underset{\sim}{\alpha}_1=0,\ \underset{\sim}{\alpha}^T\underset{\sim}{\alpha}_2 > 0\} \cup Q_3\\
&Q_1 = \{\underset{\sim}{\alpha} \,|\, |\underset{\sim}{\alpha}|=1,\ \underset{\sim}{\alpha}^T\underset{\sim}{\alpha}_1 > 0\} \cup Q_2\ .
\end{aligned}$$

Obviously, $Q_i\,(i = 1, 2, \ldots, p)$ satisfy (4.7) and

$$S_i(F) = \max_{Q_i} S(F^{\underset{\sim}{\alpha}}) = S(F^{\underset{\sim}{\alpha}_i}) = \sqrt{\lambda_i} \qquad i = 1, 2, \ldots, p\ ,$$
$$\underset{\sim}{\alpha}_i \in Q_i\ , \qquad -\underset{\sim}{\alpha}_i \notin \bar{Q}_i\ .$$

hence, $\underset{\sim}{\alpha}_i$ is the unique maximum of $S(F^{\underset{\sim}{\alpha}})$ on $\bar{Q}_i$.

For proving $s_{in} \to S(F^{\underset{\sim}{\alpha}_i})$, we have to create the conditions (1)-(4) required by Lemma 4.1. Actually, we have achieved (1) and (2):

(1) Let Ω be defined in (4.6); the continuity of $S(F_n^{\underset{\sim}{\alpha}})$ and $S(F^{\underset{\sim}{\alpha}})$ in $\underset{\sim}{\alpha}$ on Ω follows by Lemma 4.4 (1).

(2) The corollary of Lemma 4.4 has told us that

$$\lim_{n\to\infty} S(F_n^{\underset{\sim}{\alpha}}) = S(F^{\underset{\sim}{\alpha}}) \qquad P.\ F \text{ a.e.}$$

uniformly in $\underset{\sim}{\alpha}$ on Ω.

We shall verify assumptions (3) and (4).

For i=1, it is obvious: Since $S_1(F) = \max_{Q_1} S(F^{\underset{\sim}{\alpha}}) = S(F^{\underset{\sim}{\alpha}_1})$ and

$$s_{1n} = \max_{Q_1} S(F_n^{\underset{\sim}{\alpha}})\ ,$$

we need only choose $\Omega_n = \Omega_0 = Q_1$ and $P_n = I$; then by Lemmas 4.1 and 4.2, it follows immediately that

$$\left.\begin{aligned} &s_{1n} \to S(F^{\underset{\sim}{\alpha}_1}) = \sqrt{\lambda_1}\\ &\underset{\sim}{\alpha}_{1n} \to \underset{\sim}{\alpha}_1\ . \end{aligned}\right\} \qquad (n \to \infty) \qquad P.\ F \text{ a.e.}$$

Consider the case of i=2. Let P_{1n} be a rotation matrix as in Lemma 4.3 such that

$$\underset{\sim}{\alpha}_{1n} = P_{1n}\underset{\sim}{\alpha}_1$$

and define

$$Q_{2n} = P_{1n}Q_2 = \{P_{1n}\underset{\sim}{\alpha} \mid \underset{\sim}{\alpha} \in Q_2\} .$$

Obviously, Q_{2n} is a skew half of $\{\underset{\sim}{\alpha} \mid |\underset{\sim}{\alpha}|=1,\ \underset{\sim}{\alpha} \perp \underset{\sim}{\alpha}_{1n}\}$ and

$$s_{2n} = \max_{Q_{2n}} S(F_n^{\underset{\sim}{\alpha}}) .$$

Pick any maximum of $S(F_n^{\underset{\sim}{\alpha}})$ in Q_{2n}, namely $\underset{\sim}{\alpha}_{2n}$, then for obtaining

$$\left.\begin{array}{l} s_{2n} \to \sqrt{\lambda_2} \\ \underset{\sim}{\alpha}_{2n} \to \underset{\sim}{\alpha}_2 \end{array}\right. \quad (n \to \infty) \qquad P.\ \&\ a.e.$$

we need only show that

$$P_{1n} \to I \quad (n \to \infty) \qquad P\ \&\ a.s.$$

but this follow from $\underset{\sim}{\alpha}_{1n} \to \underset{\sim}{\alpha}_1$ and Lemma 4.3.

Similarly, we can find a rotation matrix P_{2n} as in Lemma 4.3 such that

$$\underset{\sim}{\alpha}_{2n} = P_{2n}P_{1n}\underset{\sim}{\alpha}_2 .$$

Define

$$Q_{3n} = P_{2n}P_{1n}Q_3 ,$$

then

$$s_{3n} = \max_{Q_{3n}} S(F_n^{\underset{\sim}{\alpha}}) = S(F_n^{\underset{\sim}{\alpha}_{3n}})$$

where $\underset{\sim}{\alpha}_{3n} \in Q_{3n}$. And $\underset{\sim}{\alpha}_{2n} \to \underset{\sim}{\alpha}_2$ gives

$$P_{2n}P_{1n} \to I \quad (n \to \infty) \qquad P.\ \&\ a.e. .$$

From Lemma 4.1 and Lemma 4.2, we have

$$\left.\begin{array}{l} s_{3n} \to \sqrt{\lambda_3} \\ \underset{\sim}{\alpha}_{3n} \to \underset{\sim}{\alpha}_3 \end{array}\right. \quad (n \to \infty) \qquad P.\ \&\ a.e. .$$

And so on. Hence, we obtain that

$$\left.\begin{array}{l} s_{in} \to \sqrt{\lambda_i} \\ \underset{\sim}{\alpha}_{in} \to \underset{\sim}{\alpha}_i \end{array}\right. \quad (n \to \infty) \qquad P.\ \&\ a.e. \qquad i = 1, 2, \ldots, p .$$

Taking the different sign of $A_i(F_n)$, $A_i(F)$ into account, we'd better write as

$$\underset{\sim}{\alpha}_{in}\underset{\sim}{\alpha}_{in}^T \to \underset{\sim}{\alpha}_i\underset{\sim}{\alpha}_i^T \quad (n \to \infty) \qquad P.\ \&\ a.e. .$$

Then it follows that

$$\Sigma_n = \sum_{i=1}^{p} s_{in}^2\underset{\sim}{\alpha}_{in}\underset{\sim}{\alpha}_{in}^T \to \Sigma \quad (n \to \infty) \qquad P.\ \&\ a.e. . \qquad \blacksquare$$

When Σ has multiple eigenvalues, we cannot make the associated eigenvectors unique anymore. But we can take another fact: any set of orthonormal basis of an eigen subspace can be the eigenvectors associated with the eigen subspace.

THEOREM 4.2.

(1) Assume that $\lambda_1 = \lambda_2 = \ldots = \lambda_p > 0$, then

$$\left.\begin{array}{lll} s_{in} \to \sqrt{\lambda_i} & i = 1, 2, \ldots, p & n \to \infty \\ \Sigma_n \to \Sigma & & n \to \infty \end{array}\right. \qquad P.\ \&\ a.e. .$$

(2) Assume that $\lambda_1 > \lambda_2 > \ldots > \lambda_k > \lambda_{k+1} = \lambda_{k+2} = \ldots = \lambda_p > 0$, then

$$s_{in} \to \sqrt{\lambda_i} \qquad (n \to \infty) \qquad \text{P. \& a.s.} \qquad i = 1, 2, \ldots, p ,$$

$$\mathbf{a}_{in}\mathbf{a}_{in}^T \to \mathbf{a}_i\mathbf{a}_i^T \qquad (n \to \infty) \qquad \text{P. \& a.s.} \qquad i = 1, 2, \ldots, k ,$$

$$\sum_{i=k+1}^{p} \mathbf{a}_{in}\mathbf{a}_{in}^T \to \sum_{i=k+1}^{p} \mathbf{a}_i\mathbf{a}_i^T \qquad (n \to \infty) \qquad \text{P. \& a.s.} ,$$

$$\Sigma_n \to \Sigma \qquad (n \to \infty) \qquad \text{P. \& a.s.} \quad .$$

<u>Proof</u>.

(1) Notice that in the case of $\lambda_1 = \lambda_2 = \ldots = \lambda_p$,

$$S(F^{\mathbf{a}}) = \sqrt{\lambda_1} \qquad \text{for any } \mathbf{a},\ |\mathbf{a}| = 1$$

and for any version of $A_i(F_n)$, namely $\mathbf{a}_{in}$ $(i = 1, 2, \ldots, p)$,

$$\sum_{i=1}^{p} \mathbf{a}_{in}\mathbf{a}_{in}^T = I = \sum_{1}^{p} \mathbf{a}_i\mathbf{a}_i^T .$$

Using Lemma 4.4 (3), we have

$$|s_{in} - \sqrt{\lambda_1}\,| = |S(F_n^{\mathbf{a}_{in}}) - S(F^{\mathbf{a}_{in}})| \to 0 \qquad (n \to \infty) \qquad \text{P. \& a.s.} \quad .$$

Hence,

$$\Sigma_n = \sum_{1}^{p} s_{in}^2\mathbf{a}_{in}\mathbf{a}_{in}^T \to \sum_{1}^{p} \lambda_i\mathbf{a}_i\mathbf{a}_i^T = \Sigma \qquad (n \to \infty) \qquad \text{P. \& a.s.} \quad .$$

(2) For $i = 1, 2, \ldots, k$, do exactly what we did in Theorem 4.1 and put

$$U_{in} = P_{in} \cdot P_{i-1n} \cdot \ldots \cdot P_{1n} \qquad (i = 1, 2, \ldots, k) \ . \qquad (4.9)$$

Then we have

$$Q_{1n} = Q_1 , \quad Q_{in} = U_{i-1n}Q_i \qquad i = 2, \ldots, k ,$$

$$s_{in} = \max_{Q_{in}} S(F_n^{\mathbf{a}}) = S(F_n^{\mathbf{a}_{in}}) \qquad i = 1, 2, \ldots, k ,$$

$$\mathbf{a}_{in} = U_{in}\mathbf{a}_i \in Q_{in} \qquad i = 1, 2, \ldots, k ;$$

and

$$\begin{matrix} s_{in} \to \sqrt{\lambda_i} , & \mathbf{a}_{in} \to \mathbf{a}_i \\ P_{in} \to I , & U_{in} \to I \end{matrix} \qquad (n \to \infty) \qquad i = 1,2,\ldots,k \qquad \text{P. \& a.s.} \qquad (4.10)$$

Put

$$Q_{k+1n} = U_{kn}Q_{k+1} .$$

We can find $\mathbf{a}_{k+1n}, \mathbf{a}_{k+2n}, \ldots, \mathbf{a}_{pn} \in Q_{k+1n}$ such that

$$s_{k+1n} = \max_{Q_{k+1n}} S(F_n^{\mathbf{a}}) = S(F_n^{\mathbf{a}_{k+1n}})$$

$$s_{k+2n} = \max_{\substack{\mathbf{a} \in Q_{k+1,n} \\ \mathbf{a} \perp \mathbf{a}_{k+1n}}} S(F_n^{\mathbf{a}}) = S(F_n^{\mathbf{a}_{k+2n}})$$

$$\cdots\cdots$$

$$s_{p-1n} = \max_{\substack{\mathbf{a} \in Q_{k+1n} \\ \mathbf{a} \perp \mathbf{a}_{k+1n}, \ldots, \mathbf{a}_{p-2n}}} S(F_n^{\mathbf{a}}) = S(F_n^{\mathbf{a}_{p-1n}})$$

$$s_{pn} = S(F_n^{\mathbf{a}_{pn}})$$

where $\mathbf{a}_{pn} \in Q_{k+1}$, $\mathbf{a}_{pn} \perp \mathbf{a}_{k+1n}, \ldots, \mathbf{a}_{p-1n}$ and $|\mathbf{a}_{pn}| = 1$. Since

$$H_{kn}^{-1}\underset{\sim}{e}_{in} \in Q_{k+1} \qquad i = k+1, \ldots, p ,$$

$$s(F^{\underset{\sim}{e}}) = \sqrt{\lambda_p} \qquad \text{any } \underset{\sim}{e} \in Q_{k+1} ,$$

$$H_{kn} \to I \qquad (n \to \infty) \qquad P. \ \& \ a.s. ,$$

we obtain, following Lemma 4.4 (3),

$$|s_{in} - \sqrt{\lambda_p}\,| = |s(F_n^{\underset{\sim}{e}_{in}}) - s(F^{H_{kn}^{-1}\underset{\sim}{e}_{in}})| \to 0$$

$$(n \to \infty) \qquad P. \ \& \ a.s. \qquad i = k+1, \ldots, p . \tag{4.11}$$

Notice that

$$\sum_{i=k+1}^{p} (H_{kn}^{-1}\underset{\sim}{e}_{in})(H_{kn}^{-1}\underset{\sim}{e}_{in})^T = \sum_{i=k+1}^{p} \underset{\sim}{e}_i\underset{\sim}{e}_i^T ,$$

we have

$$\sum_{i=k+1}^{p} \underset{\sim}{e}_{in}\underset{\sim}{e}_{in}^T - \sum_{i=k+1}^{p} \underset{\sim}{e}_i\underset{\sim}{e}_i^T = \sum_{i=k+1}^{p}\left[\underset{\sim}{e}_{in}\underset{\sim}{e}_{in}^T - (H_{kn}^{-1}\underset{\sim}{e}_{in})(H_{kn}^{-1}\underset{\sim}{e}_{in})^T\right] +$$

$$\sum_{i=k+1}^{p}\left[(H_{kn}^{-1}\underset{\sim}{e}_{in})(H_{kn}^{-1}\underset{\sim}{e}_{in})^T - \underset{\sim}{e}_i\underset{\sim}{e}_i^T\right]$$

$$= \sum_{i=k+1}^{p}\left[\underset{\sim}{e}_{in}\underset{\sim}{e}_{in}^T - (H_{kn}^{-1}\underset{\sim}{e}_{in})(H_{kn}^{-1}\underset{\sim}{e}_{in})^T\right] \to 0 \tag{4.12}$$

$$n \to \infty , \qquad P. \ \& \ a.s. .$$

Finally, (4.10), (4.11) and (4.12) together yield

$$\Sigma_n \to \Sigma \qquad (n \to \infty) \qquad P. \ \& \ a.s. \qquad \blacksquare$$

As we have seen in Theorem 4.2, when the smallest eigenvalue is the only one of multiplicity r $(r > 1)$ (including the case that all the eigenvalues are the same), it is easy to find an asymptotic identity

rotation (that is, H_{p-rn}^{-1}) which transforms all the corresponding estimates of those associated eigenvectors into that eigen subspace. But if there are some other eigenvalues of multiplicity, things will be more complicated. We discuss the case that the largest eigenvalue is the only one of multiplicity first.

THEOREM 4.3. Assume that $\lambda_1 = \lambda_2 = \ldots = \lambda_J > \lambda_{J+1} > \ldots > \lambda_p > 0$. Then

$$s_{in} \to \sqrt{\lambda_i} \qquad (n \to \infty) \qquad P. \ \& \ a.s. \qquad i = 1, \ldots, p$$

$$\sum_{i=1}^{J} \underset{\sim}{e}_{in}\underset{\sim}{e}_{in}^T \to \sum_{i=1}^{J} \underset{\sim}{e}_i\underset{\sim}{e}_i^T \qquad (n \to \infty) \qquad P. \ \& \ a.s.$$

$$\underset{\sim}{e}_{in}\underset{\sim}{e}_{in}^T \to \underset{\sim}{e}_i\underset{\sim}{e}_i^T \qquad (n \to \infty) \qquad P. \ \& \ a.s. \qquad i = J+1, \ldots, p$$

$$\Sigma_n \to \Sigma \qquad (n \to \infty) \qquad P. \ \& \ a.s. .$$

<u>Proof</u>. From the argumentation in Theorems 4.1 and 4.2, we can imagine that there will be no big problem to verify $s_{in} \to \sqrt{\lambda_i}$ $(n \to \infty)$ P. & a.s.; the difficulty lies in the first J eigenvectors.

Denote the linear subspace spanned by $A \subset R^p$ by $L(A)$ and the projection matrix onto a linear subspace L by P_L. Put

$$L_1 = L(\underset{\sim}{e}_1, \ldots, \underset{\sim}{e}_J) .$$

Notice that any orthonormal bases of L_1 can be eigenvectors of the eigen subspace L_1 corresponding to λ_1, and they give the same projection matrix onto L_1 as $(\underset{\sim}{e}_1, \ldots, \underset{\sim}{e}_J)$. Our trick is that for every n fixed, we work out a version of $A_1(F)$ in L_1 (as we will see below, that is $H_{1n}^{-1}\underset{\sim}{e}_{in}$) related to

the version Q_{1n} of $A_1(F_n)$ such that $A_1(F)$ and $A_1(F_n)$ are as close as possible ($i = 1, 2, \dots, J$).

As before, we can find $\underset{\sim}{a}_{1n} \in Q_1$ such that

$$s_{1n} = \max_{Q_1} S(F_n^{\underset{\sim}{a}}) = S(F_n^{\underset{\sim}{a}_{1n}}) ,$$

and from Lemma 4.1, we have

$$s_{1n} \to \max_{Q_1} S(F^{\underset{\sim}{a}}) = \sqrt{\lambda_1} \qquad (n \to \infty) \qquad P. \ \& \ a.s. \ . \qquad (4.13)$$

We should choose the first eigenvector in L_1 as close to $\underset{\sim}{a}_{1n}$ as possible. Thus, we pick

$$\underset{\sim}{a}_{1n}^{(1)} = P_{L_1} \underset{\sim}{a}_{1n} / |P_{L_1} \underset{\sim}{a}_{1n}| .$$

Let P_{1n} is the rotation matrix as in Lemma 4.3 such that

$$\underset{\sim}{a}_{1n} = P_{1n} \underset{\sim}{a}_{1n}^{(1)} .$$

We claim that

$$P_{1n} \to I \qquad (n \to \infty) \qquad P. \ \& \ a.s. \ .$$

If not, i.e.,

$$P\{P_{1n} \to I \ (n\to\infty)\} < 1 , \qquad (4.14)$$

then we will see a contradiction with (4.13).

Let θ_{1n} $(0 \le \theta_{1n} \le \pi/2)$ be the angle between $\underset{\sim}{a}_{1n}$ and L_1 (or equivalently, between $\underset{\sim}{a}_{1n}$ and $\underset{\sim}{a}_{1n}^{(1)}$), and $\underset{\sim}{a}_{2n}^{(1)}, \dots, \underset{\sim}{a}_{Jn}^{(1)}$ be such that $\{\underset{\sim}{a}_{in}^{(1)}, (1 \le i \le J)\}$ is a basis of L_1. Put

$$\underset{\sim}{\beta}_{1n} = (\underset{\sim}{a}_{1n} - \underset{\sim}{a}_{1n}^{(1)} \cos\theta_{1n}) / |\underset{\sim}{a}_{1n} - \underset{\sim}{a}_{1n}^{(1)} \cos\theta_{1n}| ,$$

clearly,

$$\underset{\sim}{\beta}_{1n} \perp \underset{\sim}{a}_{1n}^{(1)}, \dots, \underset{\sim}{a}_{Jn}^{(1)} ,$$

$$\underset{\sim}{a}_{1n} = \underset{\sim}{a}_{1n}^{(1)} \cos\theta_{1n} + \underset{\sim}{\beta}_{1n} \sin\theta_{1n} .$$

From (4.3) we know that (4.14) is equivalent to

$$P\{\theta_{1n} \to 0 \ (n\to\infty)\} < 1 .$$

Thus, there exists a $\theta_1 > 0$ such that

$$P\{\overline{\lim_{n\to\infty}} \theta_{1n} > 2\theta_1\} = \epsilon_1 > 0 .$$

Hence, for any sample sequence $x = (x_1, x_2, \dots) \in \{\overline{\lim_{n\to\infty}} \theta_{1n} > 2\theta_1\}$, there exists a subsequence of θ_{1n}, $\theta_{1n'}$, such that $\theta_{1n'} > \theta_1$. Then

$$\begin{aligned}
[S(F^{\underset{\sim}{a}_{1n'}})]^2 &= \underset{\sim}{a}_{1n'}^T \Sigma \underset{\sim}{a}_{1n'} \\
&= (\underset{\sim}{a}_{1n'}^{(1)} \cos\theta_{1n'} + \underset{\sim}{\beta}_{1n'} \sin\theta_{1n'})^T \Sigma (\underset{\sim}{a}_{1n'}^{(1)} \cos\theta_{1n'} + \underset{\sim}{\beta}_{1n'} \sin\theta_{1n'}) \\
&\le \lambda_1 \cos^2\theta_{1n'} + \lambda_{J+1} \sin^2\theta_{1n'} = (\lambda_1 - \lambda_{J+1}) \cos^2\theta_{1n'} + \lambda_{J+1} \\
&< \lambda_1 \cos^2\theta_1 + \lambda_{J+1} \sin^2\theta_1 .
\end{aligned}$$

This yields

$$\underline{\lim_{n\to\infty}} S(F^{\underset{\sim}{a}_{1n}}) \le \overline{\lim_{n'\to\infty}} S(F^{\underset{\sim}{a}_{1n'}}) < (\lambda_1 \cos^2\theta_1 + \lambda_{J+1} \sin^2\theta_1)^{\frac{1}{2}} < \sqrt{\lambda_1} .$$

Therefore,

$$P\{\underline{\lim_{n\to\infty}} S(F^{\underset{\sim}{a}_{1n}}) < \sqrt{\lambda_1}\} \ge P\{\overline{\lim_{n\to\infty}} \theta_{1n} > 2\theta_1\} = \epsilon_1 > 0 .$$

This contradicts (4.13). So we have proved that

$$P_{1n} \to I \qquad (n \to \infty), \qquad P. \;\&\; a.s. \;.$$

Put

$$\alpha_{in}^{(1)} = \alpha_i \qquad i = j+1, \dots, p \;.$$

Clearly, $(\alpha_{1n}, P_{1n}\alpha_{2n}^{(1)}, \dots, P_{1n}\alpha_{pn}^{(1)})$ is a basis of R^p. Define $\tilde{Q}_{2n}$ in the same way as defining Q_2 (see (4.8)), but with $\alpha_2, \dots, \alpha_p$ being replaced by $\alpha_{2n}^{(1)}, \dots, \alpha_{pn}^{(1)}$. Let

$$L_{2n} = P_{1n} L(\alpha_{2n}^{(1)}, \dots, \alpha_{jn}^{(2)})$$

$$Q_{2n} = P_{1n} \tilde{Q}_{2n} \;.$$

Obviously,

$$\alpha_{1n} = P_{1n}\alpha_{1n}^{(1)} \perp L_{2n} \;, \quad \alpha_{1n} \perp Q_{2n} \;. \tag{4.15}$$

Now we see that we have made our problem one dimension lower than before. Very similarly, we choose $\alpha_{2n} \in Q_{2n}$ such that

$$s_{2n} = \max_{Q_{2n}} S(F_n^{\alpha}) = S(F_n^{\alpha_{2n}}) \;,$$

define

$$\alpha_{2n}^{(2)} = P_{L_{2n}}\alpha_{2n} / |P_{L_{2n}}\alpha_{2n}|$$

and let:

(1) P_{2n} be a rotation matrix as in Lemma 4.3 such that
$\alpha_{2n} = P_{2n}\alpha_{2n}^{(2)}$,

(2) $\alpha_{3n}^{(2)}, \dots, \alpha_{jn}^{(2)}$ be such that $(\alpha_{2n}^{(2)}, \dots, \alpha_{jn}^{(2)})$ is a basis of L_{2n}.

Then we can show that

$$s_{2n} \to \sqrt{\lambda_1} \;, \quad P_{2n} \to I \qquad (n \to \infty) \qquad P. \;\&\; a.s. \;.$$

Actually, since $\alpha_{2n}^{(2)} \perp \alpha_{1n}$, we have

$$S(F_n^{\alpha_{2n}}) = \max_{\substack{|\alpha|=1 \\ \alpha \perp \alpha_{1n}}} S(F_n^{\alpha}) \ge S(F_n^{\alpha_{2n}^{(2)}}) \;.$$

Also, from $P_{1n}^{-1}\alpha_{2n}^{(2)} \in L(\alpha_{2n}^{(1)}, \dots, \alpha_{jn}^{(1)}) \subset L_1$ and $P_{1n} \to I \; (n \to \infty)$ P. & a.s., it follows that

$$S(F_n^{\alpha_{2n}^{(2)}}) - \sqrt{\lambda_1} = S(F_n^{\alpha_{2n}^{(2)}}) - S(F^{P_{1n}^{-1}\alpha_{2n}^{(2)}}) \to 0 \quad (n \to \infty) \quad P. \;\&\; a.s. \tag{4.16}$$

Hence, by Lemma 4.4 (3) we obtain

$$\sqrt{\lambda_1} = \max_{|\alpha|=1} S(F^{\alpha}) \ge \overline{\lim_{n\to\infty}}\, S(F^{\alpha_{2n}}) = \overline{\lim_{n\to\infty}}\, S(F_n^{\alpha_{2n}}) \ge \underline{\lim}_{n\to\infty}\, S(F^{\alpha_{2n}})$$

$$= \underline{\lim}_{n\to\infty}\, S(F_n^{\alpha_{2n}}) \ge \underline{\lim}\, S(F_n^{\alpha_{2n}^{(2)}}) = \sqrt{\lambda_1} \;. \qquad P. \;\&\; a.s.$$

This implies

$$\lim_{n\to\infty} s_{2n} = \lim_{n\to\infty} S(F_n^{\alpha_{2n}}) = \lim_{n\to\infty} S(F^{\alpha_{2n}}) = \sqrt{\lambda_1} \;. \qquad P. \;\&\; a.s. \tag{4.17}$$

We are going to show $P_{2n} \to I \; (n \to \infty)$ P. & a.s. . Let θ_{2n} be the angle between α_{2n} and L_{2n} (or $\alpha_{2n}^{(2)}$) $(0 \le \theta_{2n} \le \pi/2)$. If $\lim_{n\to\infty} P_{2n} = I$ (P. & a.s.) is not true, then by Lemma 4.3 we have

$$P(\theta_{2n} \to 0 \quad (n \to \infty)) < 1 \;. \tag{4.18}$$

Notice that (4.16) and Lemma 4.4 (3) together give

$$\lim_{n\to\infty} S(F^{a^{(2)}_{2n}}) = \sqrt{\lambda_1}\ .$$

Using this fact and by some logical argumentation like the case of P_{1n}, (4.18) will lead us to $P\{\varliminf_{n\to\infty} S(F^{a_{2n}}) < \sqrt{\lambda_1}\} > 0$. This is a contradiction with (4.17). Therefore,

$$P_{2n} \to I \quad (n \to \infty) \qquad P.\ \&\ a.s.\ .$$

In order to continue our procedure, we need to clarify two important facts:

(A) $H^{-1}_{2n}a_{1n}$ and $H^{-1}_{2n}a_{2n}$ are orthonormal vectors in L_1, where H_{in} is defined in (4.9).

(B) Put

$$a^{(2)}_{kn} = P_{1n}a_k \qquad k = J+1, \dots, p\ .$$

then $(a_{1n}, a_{2n}, P_{2n}a^{(2)}_{3n}, \dots, P_{2n}a^{(2)}_{pn})$ is an orthonormal basis of R^p. Actually, the definition of a_{2n} and $a^{(2)}_{2n}$ implies $H^{-1}_{2n}a_{2n} = P^{-1}_{1n}a^{(2)}_{2n} \in L_1$. And (4.15) implies $a_{1n} \perp a^{(2)}_{2n}$, $a_{1n} \perp a_{2n}$. This result and Lemma 4.3 together yield

$$P_{2n}a_{1n} = a_{1n}\ , \quad H^{-1}_{2n}a_{1n} = P^{-1}_{1n}a_{1n} \in L_1\ , \quad H^{-1}_{2n}a_{1n} \perp H^{-1}_{2n}a_{2n}\ .$$

Then the definition of $a^{(2)}_{2n}, \dots, a^{(2)}_{pn}$ and P_{2n} gives that $(a_{1n}, a_{2n}, P_{2n}a^{(2)}_{3n}, \dots, P_{2n}a^{(2)}_{pn})$ is a basis of R^p.

Suppose we have shown $s_{i-1n} \to \sqrt{\lambda_1}$, $P_{i-1n} \to I$ ($n \to \infty$, P. & a.s.). Similar to the case of $i = 2$, we can find $\tilde{Q}_{in}$, L_{in}, Q_{in}, a_{in}, $a^{(i)}_{kn}$ ($k = i, i+1, \dots, p$) and P_{in}, such that:

(I) $\tilde{Q}_{in}$ is constructed from $a^{(i-1)}_{i-1n}, \dots, a^{(i-1)}_{pn}$ in the way of (4.8), $L_{in} = P_{i-1n}L(a^{(i-1)}_{in}, \dots, a^{(i-1)}_{Jn})$, $Q_{in} = P_{i-1n}\tilde{Q}_{in}$ and $a^{(i)}_{kn} = H_{i-1}a_k$ ($k = J+1, \dots, p$);

(II) $a_{in} \in Q_{in}$, $s_{in} = \max_{Q_{in}} S(F^{a}_{n}) = S(F^{a_{in}}_{n})$ and $a^{(i)}_{in} = P_{L_{in}}a_{in} / |P_{L_{in}}a_{in}|$;

(III) P_{in} is a rotation matrix as in Lemma 4.3 such that $a_{in} = P_{in}a^{(i)}_{in}$, $(a^{(i)}_{in}, a^{(i)}_{i+1n}, \dots, a^{(i)}_{Jn})$ is a basis of L_{in}.

Then we can show that

(a) $s_{in} \to \sqrt{\lambda_1}$, $P_{in} \to I$, $H_{in} \to I$ ($n \to \infty$), P. & a.s.;

(b) $(H^{-1}_{in}a_{1n}, \dots, H^{-1}_{in}a_{in})$ are orthonormal vectors in L_1;

(c) $(a_{1n}, \dots, a_{in}, P_{in}a^{(i)}_{i+1n}, \dots, P_{in}a^{(i)}_{pn}) = (a_{1n}, \dots, a_{in}, P_{in}a^{(i)}_{i+1n}, \dots, P_{in}a^{(i)}_{Jn}, H_{in}a_{J+1}, \dots, H_{in}a_p)$ is a basis of R^p, for $i = 2, 3, \dots, J$.

Now we see that (a) has already provided

$$s_{in} \to \sqrt{\lambda_1} \quad (n \to \infty) \qquad P.\ \&\ a.s. \qquad i = 1, 2, \dots, J\ .$$

Using the method in (4.12), (a) and (b) together yield

$$\sum_{i=1}^{J} a_{in}a_{in}^T \to \sum_{i=1}^{J} a_i a_i^T \quad (n \to \infty) \qquad P.\ \&\ a.s.\ .$$

As for verifying of

$$\sigma_{in} \to \sqrt{\lambda_i}$$
$$\underline{a}_{in}\underline{a}_{in}^T \to \underline{a}_i\underline{a}_i^T \qquad (n \to \infty) \qquad i = J+1, \ldots, p \qquad P. \ \& \ a.s.$$

it is almost the same as in Theorem 4.1. Since when $i=J$, (c) tells us: after obtaining $\underline{a}_{1n}, \underline{a}_{2n}, \ldots, \underline{a}_{Jn}$, despite a rotation H_{Jn} (it is asymptotically identity matrix when $n \to \infty$), we are actually returning to a situation where the eigenvalues $\lambda_{J+1}, \ldots, \lambda_p$ are all different. ∎

Now we can have the general results of consistency.

THEOREM 4.4. Assume that

$$\lambda_1 = \ldots = \lambda_{I_1} > \lambda_{I_1+1} = \ldots = \lambda_{I_2} > \ldots > \lambda_{I_{t-1}+1} = \ldots = \lambda_{I_t}$$

where $I_t = P$. Then

$$\sigma_{in} \to \sqrt{\lambda_i} \quad (n \to \infty) \quad P. \ \& \ a.s. \quad i = 1, 2, \ldots, p \ .$$
$$\sum_{i=I_k+1}^{I_{k+1}} \underline{a}_{in}\underline{a}_{in}^T \to \sum_{i=I_k+1}^{I_{k+1}} \underline{a}_i\underline{a}_i^T \quad (n \to \infty) \quad P. \ \& \ a.s. \quad k = 0, \ldots, t-1, \tag{4.19}$$
$$\Sigma_n \to \Sigma \qquad (n \to \infty) \qquad P. \ \& \ a.s. \ .$$

__Proof__. For $i = 1, \ldots, I_1$, we do either exactly as in Theorem 4.1 if $I_1 = 1$, or exactly as in Theorem 4.3 if $I_1 > 1$. Then we will have

$$\begin{aligned} &\sigma_{in} \to \sqrt{\lambda_i} \\ &P_{in} \to I \qquad (i = 1, \ldots, I_1) \\ &H_{in} \to I \qquad (n \to \infty) \qquad\qquad (4.20) \\ &\sum_{i=1}^{I_1} \underline{a}_{in}\underline{a}_{in}^T \to \sum_{i=1}^{I_1} \underline{a}_i\underline{a}_i^T \ . \qquad P. \ \& \ a.s. \end{aligned}$$

For $i = I_k + 1, \ldots, I_{k+1}$ $(k = 1, 2, \ldots, t-1)$, we repeat the same procedure as the one for $i = 1, \ldots, I_1$, except we have to deal with $H_{I_k}\underline{a}_{I_k+1}, \ldots, H_{I_k}\underline{a}_p$, instead of $\underline{a}_1, \ldots, \underline{a}_p$, and everything is a little more complicated; and, of course, we have to use the results, as in (4.20), which we have obtained. ∎

Since (1.3) and $\underline{a}_{in}$ is any version of $\underline{A}_i(F)$ in a show half of the whole (p-i+1)-dimension hypersphere, (4.19) is actually

$$S_i(F_n) \to S_i(F) \qquad (i = 1, \ldots, p)$$
$$\sum_{i=I_k+1}^{I_{k+1}} \underline{A}_i(F_n)\underline{A}_i(F_n)^T \to \sum_{i=I_k+1}^{I_{k+1}} \underline{A}_i(F)\underline{A}_i(F)^T \quad k = 0, \ldots, t-1$$
$$\Sigma(F_n) \to \Sigma(F) \qquad (n \to \infty) \quad P. \ \& \ a.s. \ . \tag{4.21}$$

Now consider Huber's choice for $S(\cdot)$. If $F(\underline{x})$ is a member of an elliptic probability density family and $F_n(\underline{x})$ are the empirical distributions, then, obviously,

$$\begin{aligned} &F^{\underline{a}}(0) = F^{\underline{a}}(0_+) - F^{\underline{a}}(0) = 0 \\ & \qquad\qquad\qquad\qquad \text{for any } \underline{a} \in R^p, \ |\underline{a}| = 1 \ . \\ &F_n^{\underline{a}}(0) = F_n^{\underline{a}}(0_+) - F_n^{\underline{a}}(0) = 0 \quad a. \ s. \end{aligned}$$

Hence, according to Theorem 2.2, $S(\cdot)$ is really uniquely defined, finite and nonzero and weakly continuous almost surely on $\mathbb{F}$ ($\mathbb{F}$ is given in Lemma 4.4 (2)). This means that every condition we need for Theorem 4.1 through Theorem 4.4 is satisfied. Therefore, we have

THEOREM 4.5. If $S(\cdot)$ is Huber's scale estimator and $F(\underline{x})$ belongs to an elliptic probability density family, then the robust PP-estimates for covariance matrix and its principal components are consistent (P. & a.s.) in the sense of Theorem 4.4.

Because of consistency, we conclude that although the robust PP-estimates $A_i(F_n)$ and $\Sigma(F_n)$ may not uniquely be determined, they are asymptotically equivalent when the data come from an elliptic distribution.

5. <u>Qualitative and Quantitative Robustness</u>

Qualitative and quantitative robustness (i.e., weakly continuity and breakdown point) is discussed in this section. It is shown first that the robust PP-estimates are weakly continuous.

Assume the p-dimensional probability distribution functions $F_n(\underline{x})$ weakly converge to F_0 which belongs to an elliptic density family.

$$F_n \to F_0 \quad (w) .$$

Put

$$\tilde{\mathcal{F}} = \{F_k^{\underline{a}} \mid \underline{a} \in R^p,\ |\underline{a}| = 1,\ k = 0,1,2,\ldots\} .$$

THEOREM 5.1. Assume that $S(\cdot)$ is weakly continuous on $\tilde{\mathcal{F}}$ and that (4.19) holds. Then robust PP-estimates $S_i(F)$,

$\sum_{i=\ell_k+1}^{\ell_{k+1}} A_i(F)A_i(F)^T$ and $\Sigma(F)$ are weakly continuous at F_0.

<u>Proof</u>. Checking all the proofs of Lemma 4.1 through Theorem 4.4, with the convergence P. & a.s. replaced by ordinary convergence, we find that every step will go through except (4.5). What we need in (4.5) is that

$$d_0(F_{n_{t'}}^{\underline{a}_{t'}}, F^{\underline{a}_{t'}}) \to 0 \qquad (t' \to \infty) . \tag{5.1}$$

where d_0 metrizes the weak topology. Denote the Prohorov metric for k-dimensional probability distributions by $d_k(\cdot,\cdot)$. According to the definition of $d_k(\cdot,\cdot)$, we have

$$d_1(F_n^{\underline{a}}, F^{\underline{a}}) = \inf\{\varepsilon > 0 \mid F_n^{\underline{a}}(A) \le F^{\underline{a}}(A^{\varepsilon}) + \varepsilon \quad \text{for all } A \in \mathcal{B}_1\}$$

$$\le \inf\{\varepsilon > 0 \mid F_n(B) \le F(B^{\varepsilon}) + \varepsilon \quad \text{for all } B \in \mathcal{B}_p\}$$

$$= d_p(F_n, F) \qquad \text{for any } \underline{a} \in R^p \; |\underline{a}| = 1 . \tag{5.2}$$

where $W^{\delta} = \{\underline{x} \mid \inf_{\underline{y} \in W} |\underline{x}-\underline{y}| < \delta\}$ $(W \subset R^k)$ and $\mathcal{B}_k$ is the Borel-σ-algebra in R^k. Since the Prohorov metric metrizes the weak topology (see [10], Theorem 3.8, p. 28), (5.2) yields that

$$F_n(\underline{x}) \to F(\underline{x}) \;(W) \Rightarrow d_p(F_n, F) \to 0$$

$$\Rightarrow d_1(F_n^{\underline{a}}, F^{\underline{a}}) \to 0 \quad \text{uniformly in } \underline{a},\ |\underline{a}| = 1 .$$

Hence, (5.1) is also true in the above case. ∎

COROLLARY 5.1. If $S(\cdot)$ is Huber's choice, then $S(\cdot)$ is weakly continuous on $\tilde{\mathcal{F}}$ for any member of an elliptic density family. Hence, Theorem 5.1 holds.

Proof. According to Theorem 2.2, we have only to show that

$$F_n^{\underline{a}}(0) = F_n^{\underline{a}}(0_+) - F_n^{\underline{a}}(0) < 1 - \frac{\beta}{k^2} \quad |\underline{a}| = 1, \quad n = 0,1,2,\ldots \tag{5.3}$$

Actually, since F_0 is continuous, it is of course true that

$$F_0^{\underline{a}}(0) = F_0(\underline{0}) = 0 < 1 - \frac{\beta}{k^2} \qquad \text{for any } \underline{a},\ |\underline{a}| = 1 .$$

Also from, for example, Lemma 2.2 in [10] (see [10], p. 22), it follows that

$$0 \le \varliminf_{n\to\infty} F_n(\underline{0}) \le \varlimsup_{n\to\infty} F_n(\underline{0}) \le F_0(\underline{0}) = 0 .$$

Thus, we obtain that

$$F_n^{\underline{a}}(0) = F_n(\underline{0}) \to 0 \quad (n \to \infty) \qquad \text{for any } \underline{a},\ |\underline{a}| = 1 .$$

So, without losing generality, we can suppose that (5.3) holds for all $n = 0, 1, 2, \ldots$.

Now we show that the robust PP-estimates have much higher breakdown point than the affinely equivariant estimates, and it does not depend on the dimension.

Actually, for any p-dimensional ε-contamination model

$$F = (1-\varepsilon)F_0 + \varepsilon H ,$$

the distribution of any one-dimensional projection $\underline{a}^T\underline{x}$ is

$$F^{\underline{a}} = (1-\varepsilon)F_0^{\underline{a}} + \varepsilon H^{\underline{a}} .$$

And from the discussion at the end of Section 2, we have

(1) When $\varepsilon < \varepsilon_1$ [see (2.12)],

$$\lambda_i(F) = S(F^{\underline{a}_i}) < \infty \quad \text{for any } H, \qquad i = 1, 2, \ldots, p .$$

(Of course, $\underline{a}_i = \underline{a}_i(H)$ depends on H, $i=1,\ldots,p$.) Thus, $\hat{\Sigma}(F)$ is finite, whether the maximizing or the minimizing method is used.

If the maximizing method is used, let $\underline{a}$ be any given direction and let H put all its mass at infinity in the direction $\underline{a}$, then

$$\lambda_1(F) \ge S(F^{\underline{a}}) = \infty \qquad \text{for } \varepsilon > \varepsilon_1 .$$

Hence, $\hat{\Sigma}(F)$ is infinite. But if the minimizing method is used, it may not break down when $\varepsilon > \varepsilon_1$, since it may avoid the higher contamination direction.

(2) If F_0 is nondegenerate, then for any direction a, F_0^a is nondegenerate, too. Thus, $\varepsilon < \varepsilon_2$ implies that

$$S_i(F) = S(F^{a_i}) > 0 \quad \text{for any } H .$$

Therefore, $\hat{\Sigma}(F)$ is not singular for any H. Conversely, if $\varepsilon > \varepsilon_2$, let $H = \delta_0(x)$; then for any a, $|a| = 1$, $S(F^a) = 0$. So, $S_i(F) = 0$, $(i = 1, \ldots, p)$. $\hat{\Sigma}(F) = 0$. These are also the same for both maximizing and minimizing methods.

(3) Now assume the F_0 is degenerate in some direction β, i.e.,

$$F_0\{\beta^T x = 0\} = 1 .$$

Consider minimizing method first. If $\varepsilon < \varepsilon_3$, then for any H

$$S_p(F) = \min_{|a|=1} S(F^a) \le S(F^\beta) = 0 .$$

Obviously, if F_0 is degenerate in a k-dimensional linear subspace L ($F_0\{x \in L\} = 0$), then for any H

$$S_p(F) = S_{p-1}(F) = \ldots = S_{p-k+1}(F) = 0$$

$$a_p(F), \ldots, a_{p-k+1}(F) \in L$$

provided $\varepsilon < \varepsilon_3$. So $\hat{\Sigma}(F)$ does not misbehave at all.

The maximizing method may not pick up these phenomena, unless F_0 is degenerate into 0, i.e., $F_0\{x = 0\} = 1$. In this case

$$\varepsilon < \varepsilon_3 \Rightarrow S(F^a) = 0 \quad \text{for any } a .$$

Thus, $S_i(F) = 0$ $(i = 1, 2, \ldots, p)$ and $\hat{\Sigma}(F) = 0$ for both maximizing and minimizing methods. But it's not an interesting case.

To summarize, we conclude that robust PP-estimates have at least the same breakdown point as the projection index $S(\cdot)$ has; that is

$$\varepsilon^* = -\frac{S(0)}{\|\kappa\|}$$

except in the case where $F_0(x)$ is degenerate to a real linear subspace and the maximizing method is used. And the minimizing method may give an estimate which has a somewhat higher breakdown point than the maximizing method.

Before ending this section, there is one thing we ought to mention. Principal component analysis is a useful tool for reducing variation. But classical principal component estimation may give a totally misleading outcome when just one or two outliers occur. Robust PP-estimators do not have this problem because of their insensitivity to outliers. Especially, from the discussion in (3) above, the minimum procedure would be very helpful for detecting linear substructures of intermediate dimensions in high dimension, because it reports the degeneracy of the data honestly as long as the fraction of contamination is less than ε_3.

6. Some Comments

(1) Denote the two estimators for the covariance matrix from maximising and minimising procedures by $\hat{\Sigma}_M(F)$ and $\hat{\Sigma}_m(F)$, respectively. We can define another estimator, denoted by $\hat{\Sigma}_A(F)$, by the average of these two:

$$\hat{\Sigma}_A(F) = \frac{1}{2}\,(\hat{\Sigma}_M(F) + \hat{\Sigma}_m(F)) .$$

Since $\hat{\Sigma}_M(F)$ and $\hat{\Sigma}_m(F)$ both are consistent and weakly continuous at any member of an elliptic probability density family, obviously, so is $\hat{\Sigma}_A(F)$. And $\hat{\Sigma}_A(F)$ should also have the same equivariance and breakdown point as $\hat{\Sigma}_M(F)$, which usually has a lower breakdown point than $\hat{\Sigma}_m(F)$ has.

The simulation results show that the average procedure $\hat{\Sigma}_A(F)$ provides, on the whole, better performance than either $\hat{\Sigma}_M(F)$ or $\hat{\Sigma}_m(F)$.

(2) So far, we have not mentioned anything about correlation estimation which is also important in multivariate data analysis If we denote the elements of $\hat{\Sigma}(F)$ by $\hat{\Sigma}_{ij}(F)$, i.e.,

$$\hat{\Sigma}(F) = (\hat{\Sigma}_{ij}(F))_{p\times p} , \tag{6.1}$$

then the robust PP-estimates for corresponding correlation coefficients and correlation matrix, denoted by $r_{ij}(F)$ and $R(F)$, can be defined by rescaling, i.e.,

$$\begin{aligned} r_{ij}(F) &= \frac{\hat{\Sigma}_{ij}(F)}{[\hat{\Sigma}_{ii}(F)\hat{\Sigma}_{jj}(F)]^{1/2}} \\ R(F) &= (r_{ij}(F))_{p\times p} . \end{aligned} \tag{6.2}$$

Regarding the breakdown point, if $F(\underline{x})$ is nondegenerate, then whether $\hat{\Sigma}_{ii}(F) > 0$, in this case, is important. From the discussion in (1), (2) and (3) of Section 5, we know that the breakdown point of $r_{ij}(F)$ and $R(F)$ is

$$\varepsilon^* = \min(\varepsilon_1,\varepsilon_2,\varepsilon_3) = \min\left\{-\frac{x(0)}{\|x\|}\,,\ \frac{x(\infty)}{\|x\|} + \frac{F_0(0)}{(1-F_0(0))\|x\|}\right\}$$

for the contamination model $F = (1-\varepsilon)F_0 + \varepsilon H$.

(3) In Section 5, we mentioned that robust PP principal components themselves can do a much better job for variation reducing than the classical approach. Also, we would say that using robust PP covariance, instead of the classical one, to constitute a "robust Mahalanobis" distance would give, because of its good robustness, much better results in discrimination and clustering et al. multivariate analysis methods.

(4) It should be pointed out that the methods used here apply in some generality to the study of PP-type procedures.

Notice that Lemma 4.1 through Lemma 4.4 have nothing to do with either an elliptic probability density family or the affinely equivariance of the estimates. What we really need are the weak continuity of the projection index $S(\cdot)$ (in Lemmas 4.1, 4.2 and 4.4) and the uniqueness of the maximum points of $S(F^{a})$ on the corresponding regions (in Lemma 4.2 only). Hence, if the projection index is weakly continuous, then at least the functional for determining the first maximum value (like $S_1(F)$) should be consistent and weakly continuous. Furthermore, if the maximum points of underlying distribution (like eigenvectors $\underline{a}_i$ here) are all uniquely defined in some sense, then the functionals for searching maximum values and maximum points (like $S_i(F)$ and $\lambda_i(F)$ here) should be all consistent and weakly continuous.

7. Acknowledgments

We would like to thank Peter J. Huber for exposing us to this direction, as well as for his careful reading of the first draft and helpful comments. We also wish to express appreciation to David Donoho for his help in editing the manuscript.

References

[1] Chen, H. et al. (1974), Statistical methods for grouping corporation, Sankhya, B36, pp. 1-28.

[2] Chen, Z. and Li, G. (1981), Robust principal components and dispersion matrices via projection pursuit (manuscript).

[3] Devlin, S.J. et al. (1975), Robust estimation and outlier detection with correlation coefficients, Biometrika, 62, pp. 531-545.

[4] ―――――― (1981), Robust estimation of dispersion matrices and principal components, J. Amer. Statist. Ass. 76, pp. 354-362.

[5] Friedman, J.H. et al. (1980), Projection pursuit regression, Tech. Rep. No. 146, Dept. of Statist., Stanford Univ.

[6] Friedman, J. H. and Stuetzle, W. (1980), Projection pursuit classification (unpublished).

[7] Friedman, J. H. and Tukey, J. W. (1974), A projection pursuit algorith for exploratory data analysis, IEEE Trans. Computers, C-23, pp. 881-889.

[8] Gnanadesikan, R. and Kettenring, J. R. (1972), Robust estimates, residuals and outlier detection with multiresponse data, Biometrics, 28, pp. 81-124.

[9] Huber, P.J. (1977), Robust Covariances, in: Statistical Decision Theory and Related Topics, II, S.S. Gupta and D.S. Moore, Eds., Academic Press, New York.

[10] Huber, P.J. (1981a), Robust Statistics, John Wiley & Sons, New York.

[11] Huber, P.J. (1981b), Density estimation and projection pursuit methods (manuscript).

[12] Huber, P.J. (1981c), Projection pursuit (manuscript).

[13] Lehmann, E.L. (1959), Testing Statistical Hypotheses, John Wiley & Sons, New York.

[14] Maronna, R.A. (1976), Robust M-estimators of multivariate location and scatter, Ann. Statist., 4, pp. 51-67.

[15] Mosteller, F. and Tukey, J.W. (1977), Data Analysis and Regression, Addison-Wesley, Reading, Mass.

[16] Schönholzer, H. (1979), Robuste Kovarianz, Ph.D. Thesis, Eidgen. Technische Hochschule, Zurich.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) Robust Projection Pursuit Estimator for Dispersion Matrices and Principal Components	5. TYPE OF REPORT & PERIOD COVERED Technical Report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Guoying Li, Zhonglian Chen	8. CONTRACT OR GRANT NUMBER(s) N00014-79-C-0512, NR 042-425
9. PERFORMING ORGANIZATION NAME AND ADDRESS Harvard University, Department of Statistics, Cambridge, MA	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Code 411SP, Arlington, VA 22217	12. REPORT DATE November 1981
	13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper proposes and discusses the ROBUST PROJECTION PURSUIT ESTIMATOR for dispersion matrices and their principal components. This estimator finds robust principal components by searching, successively, for directions which maximize (minimize) a robust estimate of scale; the estimate of the dispersion matrix is constructed from the estimated principal components.

These estimators are shown below (under mild conditions) to have a number of desirable properties. They are orthogonally equivariant and, within any elliptic underlying density family, asymptotically affinely equivariant.

Furthermore, at elliptic densities, they are consistent and weakly continuous (i.e., qualitatively robust). Finally they have good quantitative robustness--their breakdown point can be as high as 1/2.

The robust projection pursuit approach is a promising alternative to other estimators of dispersion matrices.

DATE

ILME

-.8